Identification and Structural Characterisation of a Prokaryotic Bestrophin Homologue

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

René Hofer

aus

Dänemark

Promotionskommission
Prof. Dr. Raimund Dutzler (Leitung und Vorsitz)
Prof. Dr. Martin Jinek
Prof. Dr. Markus Seeger

Zürich, 2017

Acknowledgements

The work described in this thesis would not have been possible without the help of others.

First, I would like to thank Prof. Raimund Dutzler for letting me work on this project and supervising it. Thank you for all your help and support!

Also, I would like to thank Prof. Martin Jinek and Prof. Markus Seeger for serving as members of my thesis committee.

The great environment in the Dutzler-lab has contributed positively to the success of my project, so I would like to thank all lab members, past and present. Thanks also go to Novandy and Valeria for spending several years with me in L32 and to Sinem for joining the bestrophin project. I am thankful for the work done by Nathalie Cornillie as part of her master thesis project. Yvonne Neldner deserves special thanks for keeping the lab running and very special thanks for sitting next to me almost throughout the whole of my PhD.

The work involving nanobodies would not have been possible without the work of Saša Štefanić, Eric R. Geertsma and Yvonne Neldner. Beat Blattmann and Céline Stutz-Ducommun from the Protein Crystallization Center at University of Zurich have prepared many crystallisation screens for this project and have always been incredibly helpful. The Department of Biochemistry is running smoothly thanks to many people, including the IT-service and the workshop.

Of course, I am also very thankful to family and friends, some of whom are far away.

Last but not least, very special thanks to my wife Yolanda for always being there for me!

Contents

List of Figures

List of Tables

List of Abbreviations

2×YT	2× Yeast extract and tryptone (a medium for bacterial culture)
aa	Amino acid residue
ADA	N-(2-Acetamido)iminodiacetic acid
BirA	Bifunctional ligase/repressor from *Escherichia coli*
BLAST	Basic Local Alignment Search Tool
CaCC	Calcium activated chloride channel
cBEST1	Chicken bestrophin 1
CCD	Charge-coupled device
CDR1–3	Complementarity-determining region 1–3
DARPin	Designed ankyrin repeat proteins
DDM	n-Dodecyl-β-D-maltoside
DM	n-Decyl-β-D-maltoside
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
E. coli	*Escherichia coli*
ECL	Enhanced chemiluminescence
EDTA	2,2′,2″,2‴-(Ethane-1,2-diyldinitrilo)tetraacetic acid
EGTA	Ethylene glycol-bis(2-aminoethylether)-N,N,N′,N′-tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Fab	Fragment antigen binding
FSEC	Fluorescence-detection size-exclusion chromatography
FX cloning	Fragment exchange cloning
GFP	Green fluorescent protein
GluCl	Glutamate-gated chloride channel α from *Caenorhabditis elegans*
hBEST1–4	Human bestrophin 1–4
HEK293	Human embryonic kidney cells 293
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HiLiDe	Method employing high concentrations of lipid and detergent
HPLC	High-performance liquid chromatography
HRV 3C	Human Rhinovirus protease 3C
I-V curve	Current-Voltage curve
IMAC	Immobilized metal affinity chromatography
kpBEST	Bestrophin homologue from *Klebsiella pneumoniae*
LacS	Lactose permease from *Streptococcus thermophilus*
LB	Lysogeny broth (a medium for bacterial culture)
LPR	Lipid to protein ratio
MALS	Multi-angle light scattering
mBEST1–3	Mouse bestrophin 1–3
MBP	Maltose binding protein
NCBI	National Center for Biotechnology Information
Ni-NTA	Ni^{2+} chelated by 2,2′,2″-Nitrilotriacetic acid coupled to an agarose resin
OD_{600}	Optical density at 600 nm
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PelB leader	Leader sequence of pectate lyase B from *Erwinia carotovora*
PMSF	Phenylmethylsulfonyl fluoride

pS	Picosiemens
PTFE	Poly(1,1,2,2-tetrafluoroethylene)
PVDF	Poly(1,1-difluoroethylene)
RMSD	Root mean square deviation
S. spiritivorum	*Sphingobacterium spiritivorum*
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SLS	Swiss Light Source
TB	Terrific broth (a medium for bacterial culture)
TES	Tris, EDTA and sucrose containing buffer
TM	Transmembrane
UDM	n-Undecyl-β-D-maltoside
VRAC	Volume-regulated anion channel
VHH	Nanobody (variable domain of a heavy chain antibody)

The following standard three-letter and one-letter codes were used for amino acids:

Ala/A	Alanine
Arg/R	Arginine
Asn/N	Asparagine
Asp/D	Aspartate
Cys/C	Cysteine
Glu/E	Glutamate
Gln/Q	Glutamine
Gly/G	Glycine
His/H	Histidine
Ile/I	Isoleucine
Leu/L	Leucine
Lys/K	Lysine
Met/M	Methionine
Phe/F	Phenylalanine
Pro/P	Proline
Ser/S	Serine
Thr/T	Threonine
Trp/W	Tryptophan
Tyr/T	Tyrosine
Val/V	Valine

Abstract

The bestrophin family of ion channels consists of four human homologues, known as bestrophin 1–4 (BEST1–4), as well as closely related homologues found in animals and more distantly related homologues found in plants, fungi and prokaryotes. BEST1 and other animal homologues have been shown to form chloride channels that are activated by calcium binding. The currents produced by these bestrophin homologues are, however, distinct from the calcium activated chloride currents previously known to electrophysiologists, suggesting other physiological roles for bestrophins.

At the start of this project, animal bestrophins were assumed to form multimeric ion channels with several membrane spanning helices, but the molecular structure of bestrophins was not understood. It was also not known what makes bestrophins selective for certain ions, how the flow of ions is gated or how the binding of calcium causes channel opening. For prokaryotic bestrophins, even less was known, since the function of those homologues had not been studied before. It was thus not known if prokaryotic bestrophins also form anion selective channels and if they are activated by calcium, some other ligand or even by another mechanism.

By studying which functional features are conserved between the animal bestrophins and the distantly related prokaryotic homologues, I was hoping to learn more about the general principles for ion selectivity and activation in bestrophins. Since no structural information was available, I decided to attempt to solve a structure of a prokaryotic bestrophin homologue at atomic resolution and use that structure as the first step towards an understanding of the function of prokaryotic bestrophins, thereby broadening the available knowledge of the entire bestrophin protein family.

As a first step, I attempted to identify prokaryotic bestrophin homologues that could be overexpressed and purified. More than a hundred homologues were cloned from prokaryotic organisms. Homologues that could be overexpressed in *Escherichia coli* and extracted with detergent were further characterised, to assess their stability in a detergent solution. Eight homologues were found to be well expressed and sufficiently stable to allow for crystallisation screening. For seven of the homologues, crystals could be grown and their diffraction properties could be evaluated. After extensive optimisation, crystals of a bestrophin homologue cloned from *Sphingobacterium spiritivorum* were found to diffract to a resolution of 3.9 Å.

To further improve the diffraction of this homologue, nanobodies were generated. Sixteen nanobodies could be identified that formed stable complexes with the *S. spiritivorum* bestrophin homologue. Nine of these complexes were used for crystallisation trials, which yielded crystals

of seven different complexes. Eventually, one nanobody was found to significantly improve the diffraction of the *S. spiritivorum* bestrophin homologue and a complete dataset could be collected to a resolution of 3.1 Å. The structure could be solved using molecular replacement, since another bestrophin structure had in the meantime been published.

In agreement with other recently published bestrophin structures, the bestrophin homologue from *S. spiritivorum* was found to form an alpha-helical pentameric ion channel, with the individual monomers arranged symmetrically around a long, central pore axis thought to form the ion conduction pathway. Like in animal bestrophins, the gate in the *S. spiritivorum* bestrophin homologue is made up of three hydrophobic amino acid residues, indicating that the gating mechanism is at least partially conserved across evolutionary distantly related homologues. Furthermore, the structure of the ligand binding site of the *S. spiritivorum* homologue is similar to that of animal bestrophins, although the activating ligand has not yet been identified. With recent studies suggesting that prokaryotic bestrophins might form cation selective channels, structural features in the *S. spiritivorum* homologue could be identified that might contribute to the selectivity. The presented structure will thus serve as a valuable tool to direct future functional experiments.

Zusammenfassung

Die Ionenkanalfamilie der Bestrophine umfasst vier menschliche Homologe, welche als Bestrophin 1–4 (BEST1–4) bezeichnet werden. Neben den menschlichen Bestrophinen finden sich nahe Homologe auch in Tieren sowie auch entfernt verwandte Proteine in Pflanzen, Pilzen und Prokaryoten. Für BEST1 und andere tierische Homologe konnten eine Chloridkanalaktivität nachgewiesen werden, welche durch die Bindung von Kalzium aktiviert wird. Der in Funktionsstudien gemessene Strom unterscheidet sich aber von klassischen Kalzium-aktivierten Chloridströme, was eine neue physiologische Rolle der Bestrophine vermuten lässt.

Zu Beginn dieses Projektes wurde angenommen, dass tierische Bestrophine aus gleichen Untereinheiten mit mehreren Transmembranhelices bestehen. Die molekulare Struktur von Vertretern dieser Proteinfamilie war aber zu dieser Zeit noch nicht bekannt. Auch war unklar wie der Kanal bestimmte Ionen selektiert, der Ionenfluss gesteuert wird und wie die Bindung von Kalzium den Kanal öffnet. Über prokaryotische Bestrophine war noch weniger bekannt, da für diese keine Funktionsstudien vorlagen. Ebenfalls war es unklar, ob prokaryotische Bestrophine anionenselektive Kanäle formen und ob der Kanal, wie in tierischen Homologen, durch Kalzium oder mittles anderer Liganden aktiviert wird, oder ob ein völlig anderer Mechanismus der Aktivierung vorliegt.

Durch die Untersuchung, welche funktionellen Eigenschaften zwischen den tierischen und entfernt verwandten prokaryotischen Homologen konserviert sind, erhoffte ich Einblick in die generellen Prinzipien der Ionenselektivität, der Steuerung des Ionenflusses und der Aktivierung von Bestrophinen zu erlangen. Da zu Beginn dieser Arbeiten noch keine strukturellen Informationen vorhanden waren, versuchte ich die Struktur eines prokaryotischen Bestrophins aufzuklären, um damit dessen Funktion und damit die generellen Mechanismen dieser Proteinfamilie besser zu verstehen.

Der erste Schritt in dieser Arbeit war es prokaryotische Homologe der Bestrophine zu identifizieren, die überexpremiert und aufgereinigt werden konnten. Dafür habe ich über einhundert Homologe aus verschiedenen prokaryotischen Organismen kloniert und in *Escherichia coli* exprimiert. Exprimierte Homologe, die mit Hilfe von Detergentien extrahiert werden konnten, wurden weiter auf ihre Stabilität untersucht. Dabei konnten acht Proteine identifiziert werden, die sich als stabil erwiesen und die in Folge für die Proteinkristallisation verwendet wurden. Bei sieben dieser Homologe konnte ich Proteinkristalle herstellen und deren Diffraktionseigenschaften charakterisieren. Nach intensiven Optimierungen konnte Diffraktion von Proteinkristallen des Bestrophins aus *Sphingobacterium spiritivorum* bis zu einer Auflösung

von 3.9 Å gemessen werden.

Um die Diffraktion dieses Homologs zu verbessern, habe ich in Folge *Nanobodies* generiert. Sechzehn *Nanobodies* konnten identifiziert werden, die mit dem Bestrophinhomolog von *S. spiritivorum* einen stabilen Komplex bildeten. Mit neun verschiedenen Komplexen wurden Kristallisationsexperimente durchgeführt, wobei für sieben Komplexe Proteinkristalle erhalten wurden. Unter diesen verschiedenen Kristallen konnte ich einen *Nanobody* identifizieren, der die Diffraktion des *S. spiritivorum*-Homologs bedeutend verbesserte. Mit Kristallen dieses Komplexes konnte ich einen kompletten Datensatz mit einer Auflösung von 3.1 Å erhalten. Die Kristallstruktur des Komplexes konnte ich mittels 'Molecular Replacements' mit einer Struktur eines anderen Bestrophins, das in der Zwischenzeit von einer anderen Gruppe publiziert wurde, bestimmen.

In Übereinstimmung mit der in letzter Zeit veröffentlichten Strukturen konnte gezeigt werden, dass das Bestrophinhomolog von *S. spiritivorum* einen alpha-helikalen pentameren Ionenkanal bildet, wobei die einzelnen Monomere symmetrisch um eine Achse angeordnet sind. Die Pore für den Ionenfluss befindet sich entlang der Symmetrieachse. Wie in den tierischen Bestrophinen wird der Ionenfluss im *S. spiritivorum*-Homolog durch drei hydrophobe Aminosäurereste an der Engstelle des Kanals in der Mitte der Membran beeinflusst. Dies deutet darauf hin, dass der Ionenfluss bei den evolutionär entfernt verwandten Homologen ähnlich reguliert wird. Ausser der Pore ist auch die Struktur der Ligandenbindungsstelle im *S. spiritivorum*-Homolog ähnlich zu den tierischen Proteinen, wobei für prokaryotische Bestrophine noch kein Ligand identifiziert werden konnte. Kürzlich veröffentlichte Studien legen nahe, dass prokaryotische Bestrophine kationenselektive Kanäle formen. In der vorliegenden Struktur des *S. spiritivorum*-Homologs konnten strukturelle Merkmale identifiziert werden, die einen Einfluss auf die Selektivität haben könnten. Die in dieser Doktorarbeit präsentierte Struktur bildet deswegen eine wertvolle Grundlage für zukünftige funktionelle Experimente.

1. Introduction

1.1. Ion flow across membranes

All living cells are surrounded by a cell membrane, which separates the outside from the inside of the cell. In eukaryotic cells, membranes also enclose intracellular compartments, separating the inside of those compartments from the cytosol, but in the following only the cell membrane is discussed.

The cell membrane mainly consists of lipids and proteins. The lipid molecules are the main constituents of the cell membrane. They spontaneously form lipid bilayers, which act as barriers. The core of the lipid bilayer is very hydrophobic, making it difficult for many water-soluble compounds to diffuse across the membrane. Larger, polar molecules permeate only slowly across a lipid bilayer, while ions are essentially impermeant.

The cell membrane is, unlike a pure lipid bilayer, not impermeant to ions and many larger polar molecules. It is in fact semi-permeable, thanks to the presence of many kinds of membrane embedded proteins that allow solutes to move across the membrane. Many other kinds of proteins are also present in the membrane, but in this thesis the focus will be put on proteins that are directly involved in the movement of ions across the membrane.

1.1.1. Ion channels

By the action of membrane proteins, a living cell maintains different concentrations of various ions on the inside compared to the outside, which essentially creates a chemical gradient across the membrane. Due to the selective permeability properties of biological membranes, the downhill flow of certain ions generates an electrical potential difference across the membrane, also known as the membrane potential. The combined effects of the membrane potential and the chemical gradient can be described as an electrochemical potential. The flow of any ion across the membrane is affected by this electrochemical potential. For ions flowing downhill, all that is needed is an open path that permits the ions to pass the hydrophobic barrier of the membrane. However, for ions to move in the opposite direction, an open path is not enough and the input of energy is required. Not surprisingly, proteins that allow ions to move across the membrane can be divided into two main groups. Ion channels allow the passive flow of ions down the gradient, whereas ion transporters move ions against this gradient. Ion transporters will not be discussed further, since the topic of this thesis concerns an ion channel.

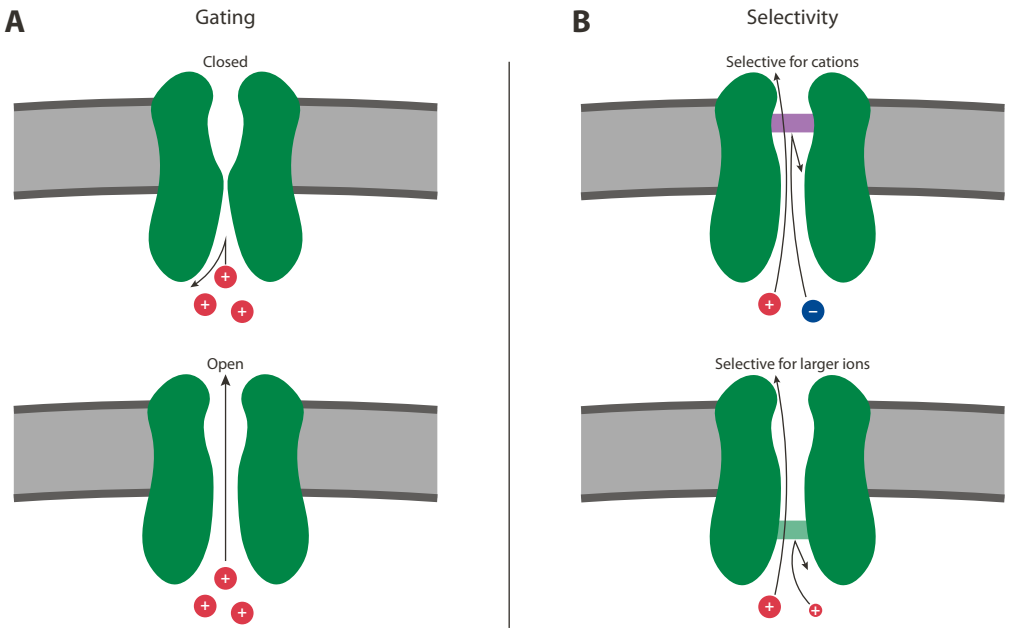


Figure 1 – Gating and selectivity of ion channels. Channels are show in green, cell membranes in grey. **(A)** A gated channel shown in closed (upper panel) and open (lower panel) conformations. Only the open state allows ions to flow through the channel. **(B)** Examples of selectivity in ion channels. The upper panel shows a cation selective channel with a selectivity filter (light purple) that only permits the flow of cations. The lower panel shows a channel that is selective for larger cations. Smaller cations cannot flow through the selectivity filter (light green).

If the cell membrane would only contain permanently open ion channels of different selectivities, it would be difficult for a cell to maintain an electrochemical gradient. For this reason, channels typically contain a gate that can be opened or closed to regulate the flow through the channel. The use of the word gate is conceptual, since the actual mechanism of gating varies between different families of ion channels. Generally gating can be understood as a transition that occurs as the channel changes from a closed to an open state or the other way round (Figure 1a).

Another characteristic of ion channels concerns their selectivity, which means that only certain ions are permitted to flow through a typical channel (Figure 1b). The region that is responsible for the selectivity is often referred to as the selectivity filter. Many ion channels are charge selective, i.e. they are selective for either anions or cations. Additionally, some ion channels are also size-selective, meaning that they only allow ions of a certain size to pass through the pore. As an example, potassium channels show a strong selectivity for potassium ions over sodium ions [1].

1.1.2. Activation of ion channels

A typical ion channel containing a gate needs to be activated in some way before the gate opens. Ion channels can be grouped into different categories, based on the mechanism of activation. Some commonly observed activation mechanisms are summarised in figure 2 and described briefly below.

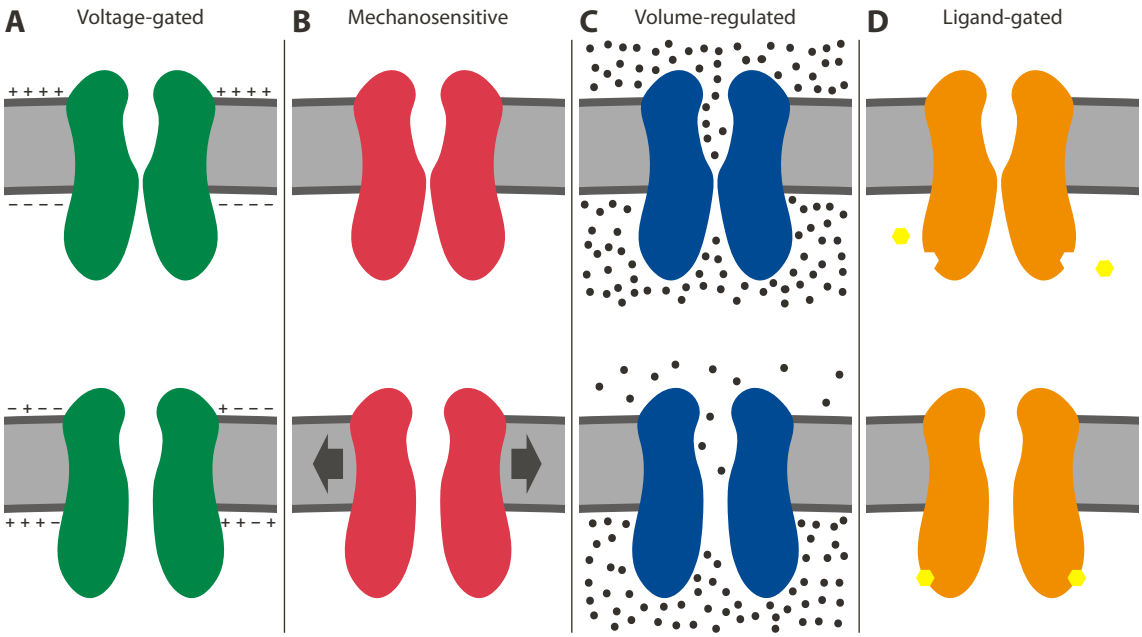


Figure 2 – Examples of channel activation mechanisms. Cell membranes are shown in grey. The upper panels show the closed states, whereas the lower panels show the open states. **(A)** A voltage-gated channel (green) becomes activated as the membrane depolarises. **(B)** A mechanosensitive channel (red) opens as a response to mechanical stretch of the membrane. **(C)** A simplified mechanism illustrating how a volume-regulated channel (blue) responds to a drop in osmolarity on the outside of the cell. Black dots indicate solutes. **(D)** A ligand-gated channel (orange) is activated by the binding of a ligand (yellow).

In some channels, known as voltage-gated ion channels, the gating is regulated by the membrane potential. These include for example voltage-gated sodium channels that are involved in the propagation of action potentials in neurons [2].

Another regulatory mechanism is seen in the mechanosensitive channels, which are regulated by mechanical pressure that induces stretch or curvature in the membrane. The best studied mechanosensitive channels are prokaryotic homologues [3], but eukaryotic mechanosensitive channels also exist [4].

Other channels, the so-called volume regulated channels, respond to changes in cell volume. Some volume regulated channels might respond to changes in membrane stretch or curvature or respond to forces exerted by intracellular or extracellular proteins [5]. However, for the volume regulated anion channel, VRAC, the underlying mechanism appears to be the sensing of a change in the ionic strength, which occurs as an effect of hypotonic stress. [6,7].

Finally, many channels are found to be activated by a ligand. Examples include various receptors of the Cys-loop family that are activated by different neurotransmitters [8], calcium activated potassium channels [9], or calcium activated chloride channels, as discussed later.

Importantly, one type of channel might also be regulated by several different mechanisms at the same time. For instance, the large conductance calcium activated potassium channels are not only regulated by calcium, but also by voltage [10]. Additionally, other activation mech-

anisms exist, such as activation by other proteins, e.g. by G proteins [11], or activation by a change in pH [12].

1.2. Calcium activated chloride channels

The study of membrane permeability by electrophysiology dates back many decades to a time where the molecular identities of ion channels were not yet known [1]. This was also the case for calcium activated chloride currents that were discovered long before the molecular identity of calcium activated chloride channels (CaCC) could be determined. Already in 1982, anion currents that required influx of calcium and could be blocked by intracellular EGTA were observed in retinal rod cells from the tiger salamander (*Ambystoma tigrinum*) [13]. The influx of calcium was achieved by depolarisation of the membrane, which opened voltage-dependant calcium channels. The observed anion currents were described as outward currents, which by electrophysiological convention is equal to either an influx of negatively charged ions or efflux of positively charged ions. Since caesium had been added to the inside of the cell to block the efflux of potassium, the observed currents were believed to be due to the influx of anions, presumably chloride.

Currents with similar properties were shortly after also observed in *Xenopus laevis* oocytes [14]. Here the currents were found to be increased when the extracellular concentration of calcium was increased and abolished when calcium current blockers such as magnesium were added to the extracellular solution. All of this was consistent with the idea that the influx of calcium activated the observed currents. While it was found to be necessary to depolarise the membrane to above −20 mV, consistent with the activation of inward calcium currents, it was also found that making the membrane potential too positive essentially abolished the observed anion current, consistent with calcium not flowing into the cell, when the inside was too positive. Furthermore, the observed anion current was found to be transient, since it decreased over time. Similar currents were also observed in a study of cells isolated from rat tear glands [15] and in a study with *Xenopus laevis* oocytes [16], which confirmed that the observed currents were indeed chloride currents. A later study conducted on cells isolated from rat tear glands, verified the calcium dependence in a more direct manner, by varying the intracellular calcium concentration. [17].

Since those early studies, currents from calcium activated chloride channels have been observed in many different cell types and they appear to be involved in many physiological processes, including, but not limited to, taste and olfactory transduction, phototransduction, regulation of neuronal and cardiac excitability, smooth muscle contraction and fluid secretion in epithelia and glands [18].

Certain characteristics of the current arising from these channels have over the years been established without knowing the molecular identity of the channels. At low intracellular calcium concentrations, the observed current is voltage dependent and the activation is time dependant. However, those dependencies disappear at concentrations above approximately 1 μM of Ca^{2+} [19]. The time dependence at low calcium concentrations can be observed as a slow activation, while the voltage dependence can be observed as an outward rectification (higher currents at positive voltages).

Due to the high concentrations of chloride ions present in living organisms, the name calcium activated chloride channel seems reasonable. CaCCs are, however, rather non-selective, allowing anions of different sizes to permeate through the channel. Larger anions have been found to have a higher permeability following the sequence $I^- > NO_3^- > Br^- > Cl^- > F^-$ [1].

1.2.1. TMEM16 and bestrophins

As previously mentioned, many early studies of CaCCs were done without knowing the molecular identities of the channels. Although different candidate proteins were suggested over the years, it was only in the 2000s that anoctamins and bestrophins were proposed to form CaCCs.

The first member of the bestrophin protein family was identified as the gene responsible for an eye disease called Best Vitelliform Macular Dystrophy [20]. The encoded protein, now known as BEST1 (hBEST1 when referring to the human homologue), was suggested to be a transmembrane protein with at least four transmembrane segments, based on hydropathy plots. Later, three additional bestrophin homologues were identified in the human genome. [21]. These are now referred to as BEST2–4 (or hBEST2–4 for the human homologues). hBEST1 and hBEST2 were shown to form chloride channels and hBEST1 was furthermore shown to be activated by sub-micromolar concentrations of calcium [22]. However, unlike the classical CaCCs described in the previous sections, hBEST1 and hBEST2 did not show any time dependency of activation and despite using sub-micromolar concentrations of calcium, no strong voltage dependence was observed [22]. In a later study, the homologues hBEST3 and hBEST4 also did not show the characteristics expected for classical CaCCs [23]. hBEST3 did show strong rectification and time dependence of activation, but unlike classical CaCCs, the rectification was in the inward direction and the currents continuously increased over the course of several seconds.

While the members of the bestrophin protein family do appear to form chloride channels that are activated by intracellular calcium, their currents do not appear to match the currents that have typically been described as CaCC currents. Instead, those classical CaCCs now appear to have been identified as members of the TMEM16 (or anoctamin) protein family [24–26]. Two

members of this family, TMEM16A and TMEM16B appear to form calcium activated chloride channels with properties that are in agreement with the classical CaCCs [27].

1.3. Closer look at the bestrophin protein family

Members of the bestrophin family are thus not responsible for the classical calcium activated chloride currents, but they still appear to form chloride channels that are activated by calcium and they still appear to be involved in physiologically relevant processes as it will be described later.

As previously mentioned, four bestrophin homologues are found in the human genome (hBEST1–4), but only three (mBEST1–3) exist in mouse. Typically, 3–4 homologues are found in all vertebrates. All vertebrate bestrophins can phylogenetically be divided into four groups, each containing one of the four human homologues [28]. For non-vertebrate animals, the number varies significantly. Based on a recent search of the UniProt database, four homologues are found in *Drosophila melanogaster*, while twenty-six are found in *Caenorhabditis elegans*. Additionally, distantly related homologues are also found in plants, algae, fungi [28] and in prokaryotes [29].

The four human bestrophins, and other animal bestrophins, vary greatly in length. hBEST1–4 consist of 585, 509, 668 and 473 amino acid residues, respectively. Several splice variants have been reported for hBEST1 [20] and hBEST3 [21], but it is unknown if these are of any physiological importance. Chicken BEST1 (cBEST1), which is discussed later, consists of 762 amino acid residues in the isoform which is most comparable to canonical hBEST1 sequence. As shown in figure 3, there is a high degree of sequence conservation between different bestrophins until approximately residue number 370 (numbering based on hBEST1). The following C-terminal region appears poorly conserved and is also aligning poorly when comparing different homologues. Furthermore, large parts of the non-conserved C-terminal regions are predicted to be intrinsically disordered, based on meta analyses of the human bestrophin homologues with the metaPrDOS disorder prediction web server [30]. Such intrinsically disordered regions are often observed in the cytoplasmic domains of eukaryotic plasma membrane proteins [31].

After the discovery of the bestrophin protein family, two studies tried to predict the transmembrane topology of these proteins. In a study by Tsunenari et al. [23], insertion of TEV cleavage sites and N-linked glycosylation sites was used along with MTSET modifications, to propose a model of hBEST1 with four membrane spanning helices and a re-entrant loop, where the N- and C-termini were located on the inside of the cell. This was in partial agreement with a hydropathy analysis showing two pairs of strongly hydrophobic peaks, interspersed with two weakly hydrophobic peaks. However, when comparing to bacterial sequences, only the

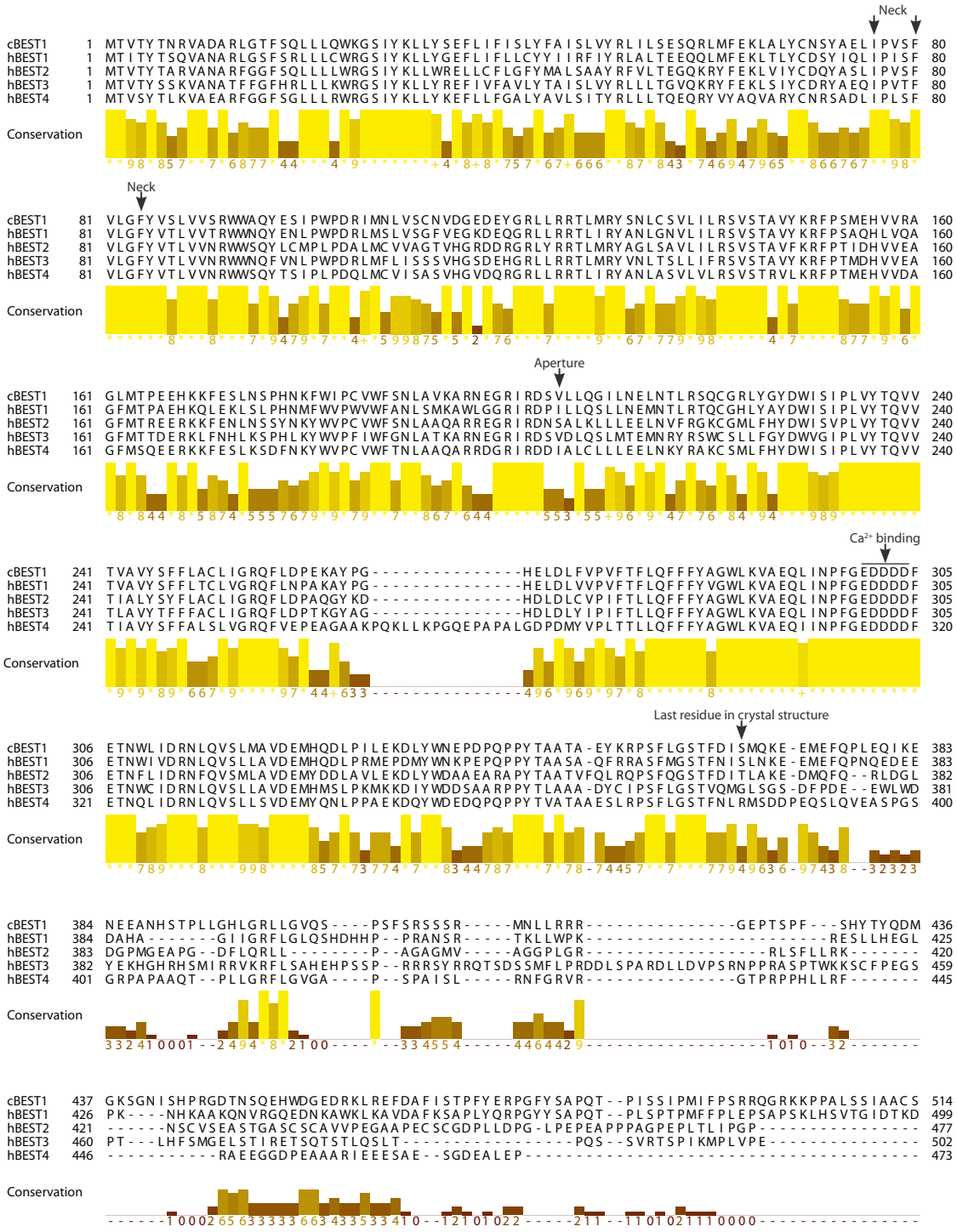


Figure 3 – Multiple sequence alignment of cBEST1 and hBEST1–4. Only parts of the disordered and non-conserved C-termini are shown. The constricting residues in the neck and aperture and the conserved Ca^{2+} binding site are labelled.

four strongly hydrophobic peaks seem to be conserved [29], suggesting a potential flaw of the Tsunenari model. Another study by Milenkovic et al. [32] analysed the ability of putative membrane spanning helices to insert into the endoplasmic reticulum membrane. Based on this, a model with four membrane spanning helices was proposed. Between the first and the second helices as well as between the third and fourth, short extracellular loops were pre-

dicted, while a large intracellular segment of more than hundred amino acid residues was predicted between the second and third transmembrane helices. The N- and C-termini were predicted to be on the cytosolic side. Crystal structures (described in a following section) have since then confirmed the model proposed by Milenkovic et al., suggesting that this probably represents the general topology of bestrophin homologues.

Regarding the oligomeric assembly of bestrophins there has been some confusion. One study suggested that hBEST1 forms a tetramer or pentamer [22], while another study suggested a dimer [33]. With the availability of pentameric crystal structures (described later), a pentameric assembly now seems to be a general feature of the family that is necessary for the function of the protein. It has been shown by co-transfection that hBEST1 can associate with hBEST2 [22], suggesting that heteropentamers might form. The physiological relevance is, however, unknown, since only the function of homopentamers has been studied.

Like many other chloride channels, bestrophins appear to be permeable to several different anions. For hBEST1 the larger ions appear to permeate more easily through the channel, according to the following typically reported permeability sequence $SCN^- > NO_3^- > I^- > Br^- > Cl^- > F^-$ [28]. Bestrophins have additionally been shown to be permeable to HCO_3^-, with a permeability that is slightly lower than for Cl^- [34]. Furthermore, mBEST1 has been reported to be permeable to gluconate [35] and glutamate [36], although that could not be verified for cBEST1, neither using a liposome assay [37] nor in bilayer experiments [38].

1.3.1. Activation by calcium

As previously mentioned, hBEST1 has been shown to be activated by calcium [23]. Calcium activation has also been shown for mBEST2 [39] and hBEST4 [40]. EC50 values of 141 nM Ca^{2+} for hBEST1 [41], 230 nM for mBEST2 [39] and 230 nM for hBEST4 [40] have been reported. While hBEST2 has been reported to mediate currents in the presence of calcium [22], the calcium dependence has not been investigated. It would, however, not be surprising if it is, similar to mBEST2, inactive in the absence of calcium.

Initially, it was not clear if bestrophins were activated directly by binding of calcium or by interacting with other calcium binding proteins. Sequence analysis did suggest a possible calcium binding site with the conserved sequence Glu-Asp-Asp-Asp-Asp, located immediately after the last transmembrane helix (Figure 3), since a similar stretch of acidic residues has been found to form the calcium binding bowl in large conductance calcium activated potassium channels [42]. Mutations of these residues in hBEST1 resulted in non-functional protein, without affecting the plasma membrane localisation, but mutations in residues following this acidic stretch also affected the activation by calcium [41]. Later structural studies on cBEST1 have, however, confirmed that calcium can be directly bound by the acidic residues in this re-

gion [37]. Furthermore, since purified and reconstituted cBEST1 can be activated by calcium, it can be concluded that the activation is due to direct binding of calcium and not due to other endogenous factors [37,38].

1.3.2. Possible regulation by phosphorylation

While it seems clear that calcium can activate some vertebrate bestrophins, it is still possible that other factors might be involved in the activation as well. Marmorstein et al. [43] noticed that porcine BEST1 coimmunoprecipitated with protein phosphatase 2A, when it was extracted from porcine retinal pigment epithelium cells. Furthermore, hBEST1 was shown to be phosphorylated when expressed in cultured retinal pigment epithelium cells. The level of phosphorylation was, however, not quantified and the functional state of the phosphorylated bestrophin was not confirmed. Thus, the study gives little more than an indication that phosphorylation of porcine BEST1 and hBEST1 might be taking place. In another study Xiao et al. [44] identified a phosphorylation site in hBEST1 for protein kinase C (Ser358), which appeared to affect the function of hBEST1. The same group had previously demonstrated a rundown of the hBEST1 current, which became faster in the presence of higher concentrations of calcium [41]. When phosphorylated or when a glutamate residue was introduced at this position, the rundown was decreased, meaning that the channel remained active for longer.

Interestingly, in a study on fruit fly BEST1, which activates slowly after the addition of calcium, ATP was found to increase the rate of activation, while a non-hydrolysable ATP analogue decreased the activation rate [45]. The same study found that inhibitors of calcium/calmodulin-dependent protein kinase II decreased the current amplitude. Together these findings suggested another mechanism for calcium activation of a bestrophin besides the direct calcium binding to the described site, namely through a calcium dependent kinase. As the calcium binding site is conserved in fruit fly BEST1 (dBEST1), it is possible that these distinct mechanisms could operate in parallel. Although the role of phosphorylation appears to be different for dBEST1 compared to hBEST1, it does seem to affect the activity in both cases, suggesting that phosphorylation might generally play a role in the modulation of bestrophin activity, although the exact effect might vary between different homologues and possibly between different cell types.

1.3.3. Regulation by a C-terminal motif

hBEST3 and mBEST3 have been reported to be very slowly activated, mediating inward rectifying currents only after several seconds [23,46]. This is unlike the other human and murine bestrophins that are activated in a largely time-independent manner, without any rectification. This has led to the notion that hBEST3 and mBEST3 may not be activated by calcium, even though they might bind calcium and require it for their function.

As previously mentioned, most animal bestrophins have a rather long C-terminus that is predicted to be largely unstructured. In mBEST3 it was discovered that deletion of most of the C-terminus, starting from residue 353, led to larger currents with time-independent activation and no rectification [46]. Since deletions after residue 405 showed effectively the same behaviour as the wild-type mBEST3, the responsible region could eventually be pinpointed to a small motif consisting of residues 356–362 and it could furthermore be shown that the deletion of this motif in hBEST3, had the same effect as the deletion in mBEST3. The motif had the sequence Ile-Pro-Ser-Phe-Leu-Gly-Ser and mutation of any residue, except proline, to alanine activated the channel [47]. A similar sequence was found in mBEST2 (Gln-Pro-Ser-Phe-Gln-Gly-Ser), but when the sequence of this motif in mBEST2 was mutated to that of mBEST3, the currents of mBEST2 were increased, suggesting that this region also had an inhibitory role in mBEST2 [47]. This has led to the discovery that some feature of mBEST2 in the region 405–454 might suppress the effect of the inhibitory sequence. Consequently, introducing the 405–454 region from mBEST2 into mBEST3, increased the current of mBEST3, whereas introducing the same region from mBEST3 into mBEST2, decreased the currents of mBEST2. The authors thus hypothesised that the effect of the inhibitory sequence was outcompeted by the effect of the facilitatory region (405–454) in mBEST2 [47]. Unfortunately, it was not investigated further which exact residues might be responsible for this facilitatory effect.

1.3.4. Bestrophins and volume regulation

Another regulatory mechanism that has been proposed for certain bestrophin homologues, is volume sensitivity. Chien and Hartzell [48] found that osmotically activated anion currents in *Drosophila* S2 cells, could be abolished with RNA interference directed against the endogenous dBEST1 and rescued by transfecting the cells with a plasmid encoding dBEST1. Calcium was not found to be required for this osmotically activated current, but it did increase the measured peak current. Calcium could also activate the current alone, suggesting a dual mechanism of activation, in addition to the previously discussed possible effects of phosphorylation. A later genome wide RNA interference study by Stotz and Clapham [49], independently found that dBEST1 was responsible for anion currents activated by extracellular hypo-osmotic solutions in *Drosophila* S2 cells. In the same study, an mBEST2 chimera, where the first 64 residues from the dBEST1 N-terminus replaced the native N-terminus of mBEST2, was found to be activated by hypo-osmotic solutions, unlike the wild-type mBEST2.

However, a later study by Chien and Hartzell did not find any differences in volume-regulated anion currents observed in peritoneal macrophages isolated from wild-type or BEST1$^{-/-}$/BEST2$^{-/-}$ mice [50], suggesting that bestrophins probably are not responsible for the classical VRAC currents observed in mammalian cells. Considering the recent discovery that VRACs are formed by LRRC8 proteins [7,51], this does not seem surprising. Nonetheless, the pos-

sibility remains that bestrophins in certain cell types could be either directly responsible for VRAC currents, in some way be involved in the generation of those currents or be regulated by changes in osmolarity.

The notion that bestrophins might be involved in volume regulation in some cell types has recently gained more credibility following a study by Milenkovic et al. [52]. The study has characterised VRAC currents in retinal pigment epithelium cells derived from human induced pluripotent stem cells. These currents were not affected by knockdown of LRRC8A, but were instead greatly reduced in the cells from patients with pathological BEST1 mutations. Furthermore, a greatly reduced fertility was observed in male $BEST1^{-/-}$ mice, consistent with a high BEST1 expression level in the testes of wild-type mice. Sperm cells from the $BEST1^{-/-}$ mice showed lower motility and abnormal morphology. These sperm cells were found to cope less well with changes in osmolarity, which indirectly suggested a role for BEST1 in volume regulation. Furthermore, both hBEST1 and mBEST1 were tested in a *Xenopus laevis* oocyte swelling assay. Oocytes injected with only aquaporin-1 mRNA were quickly rupturing when exposed to a hypotonic solution, but this effect could be significantly delayed by co-expressing hBEST1 or mBEST1. Swelling induced currents were more than twice as high when oocytes were co-expressing hBEST1 or mBEST1 along with aquaporin-1, compared to aquaporin-1 alone.

The findings by Milenkovic et al. are, however, somehow contradicted by an earlier study by Fischmeister and Hartzell [53] that investigated hBEST1 and mBEST2. Overexpression of either of those proteins in HEK293 cells did not appear to have any significant effect on the currents induced by hypotonic solutions. Milenkovic et al. argued that HEK293 cells are not ideal for studying BEST1 in this case, since a large fraction of overexpressed BEST1 appears to be retained inside the cells. Also, the use of mBEST2 by Fischmeister and Hartzell might not have been ideal, since the properties of mBEST1 in this regard appear to be more directly comparable to hBEST1. Based on the study by Milenkovic et al., it does seem likely that BEST1 is involved in the generation of VRAC currents in certain cell types, but more studies are needed to fully understand this function of BEST1.

1.3.5. Potential interactions with other proteins

Various reports have suggested that Best1 might act as a regulator of voltage-dependant calcium channels ($Ca_V1.3$). Rosenthal et al. [54] observed a faster activation of currents which were believed to be mediated by L-type calcium channels in rat retinal pigment epithelium cells, when these were transfected with a hBEST1 plasmid. The observed effects, however, appear relatively small and one might also question the relevance of examining hBEST1 in cultured cells from rats.

Two later studies [55,56] from the same group have found similar effects in Chinese hamster ovary cells that were microinjected with the subunits of a voltage-dependant calcium channel from rats and human BEST1. These studies argued that there was an interaction between the so-called β-subunit of the rat calcium channel and proline-rich motifs in the C-terminus of human BEST1. The evidence for this interaction was membrane colocalisation and coimmunoprecipitation. The latter, however, requires detergent extraction of the proteins and as many membrane proteins are not stable in detergents, such results should be interpreted with caution. Additionally, misfolded protein retained in the ER could potentially also be extracted by the used detergents, which would further complicate the interpretation of the results.

Another group has reported that the current from a rat L-type voltage-dependent calcium channel ($Ca_V1.3$) expressed in HEK293 cells was inhibited when human BEST1 was coexpressed [57]. No apparent effect was observed when mBEST1, mBEST2 or mBEST3 were used instead of hBEST1. The inhibitory effect could also be observed by coexpression of a C-terminal hBEST1 fragment that did not have any channel activity. By coexpressing various truncated fragments of hBEST1 the observed effect was located to the region 330–370. The biochemical stability of the truncated proteins was unfortunately not reported. A direct interaction between a C-terminal fragment of hBEST1 and a β-subunit (β_{2a}) was shown by coimmunoprecipitation, but here the results should probably be interpreted with the same caution as described above. The interaction was furthermore analysed by a pull-down assay of a C-terminal hBEST1 fragment and a rat β-subunit (β_{2a}). Both were, however, expressed in *Escherichia coli* and their biochemical stability was not evaluated, thus also these results should be interpreted with caution.

Based on the described studies, it is a possibility that an interaction can take place between human BEST1 and voltage-dependant calcium channels from rats. It is, however, still unclear if the same interaction would be seen with voltage-dependant calcium channels from humans. Also, due to the discussed weaknesses, the conclusions presented in the described studies should be interpreted with caution.

1.3.6. Bestrophin-1 in the brain

Several studies from one group have recently suggested that BEST1 might have a physiological role in the brain. Park et al. [36] measured calcium activated anion currents in cultured mouse astrocytes and found that those currents could be reduced by a knock-down of mBEST1, which appeared to be expressed at high levels [36,58]. In another study mBEST1 was suggested to be responsible for GABA release from glial cells in the cerebellum [59]. It has also been suggested that two different modes for release of glutamate in astrocytes, fast and slow release, are mediated by TREK-1 and BEST1, respectively [60]. The proposed involvement of TREK-1 is surprising, since TREK-1 is a potassium channel. Potassium channels are generally

very selective for potassium and the permeation of large, negatively charged ions through TREK-1 is thus a controversial proposal. As previously mentioned, glutamate could also not be found to permeate through cBEST1 [37,38], making the proposal of a potential permeation through mBEST1 controversial as well.

The role of BEST1 in the brain is an intriguing topic. BEST1 does seem to be expressed in certain brain cells [36,58,61], but the question whether its function is the release of GABA and glutamate requires further investigations.

1.3.7. Other physiological roles of bestrophins

Since the BEST1 gene was discovered as the gene responsible for Best vitelliform macular dystrophy, a large focus has been put on the investigation of possible roles of BEST1 in ocular epithelia. The exact role of BEST1 in the eye is not entirely clear, but mutations in the BEST1 gene are known to cause several different eye diseases [62].

mBEST2 has been studied in various contexts. It has been proposed to secrete HCO_3^- in goblet cells of the colon [63] and to be involved in sweat secretion [64]. It has also been proposed to form CaCCs in olfactory sensory neurons [65], but a later study found the same currents in BEST2$^{-/-}$ mice, so probably BEST2 has some different role in that type of tissue. Unfortunately, several studies on bestrophins were conducted before members of the TMEM16 family of proteins were shown to form the classical CaCCs. It is thus possible that some of the physiological functions that were initially assigned to bestrophins, might instead be mediated by members of the TMEM16 family.

1.3.8. Bestrophins in non-animal organisms

Although bestrophin homologues can be found in other eukaryotes apart from animals as well as prokaryotes, few studies have investigated the properties of those distantly related homologues. One study found that a homologue from the filamentous fungus *Aspergillus nidulans* formed an anion channel, that was apparently permeable to citrate [66]. Interestingly, the fungal bestrophin homologue required calcium for activation, even though only two out of five residues were conserved in the previously described acidic stretch, corresponding to Asp301 and Asp304 in hBEST1. The EC50 was reported to be 1.6 μM Ca^{2+}, which is considerably higher than the values reported for animal bestrophins. As previously mentioned, an EC50 value of 141 nM Ca^{2+} has been reported for hBEST1 [41].

1.4. Bestrophin structures

During the course of the project described in this thesis, two bestrophin structures were solved independently by two different groups. Yang et al. [67] solved the structure of the prokaryotic

homologue kpBEST at a resolution of 2.3 Å. Later Kane Dickson et al. [37] solved the structure of cBEST1 at 2.85 Å resolution. As described below, these two homologues were shown to form pentameric channels of a similar architecture.

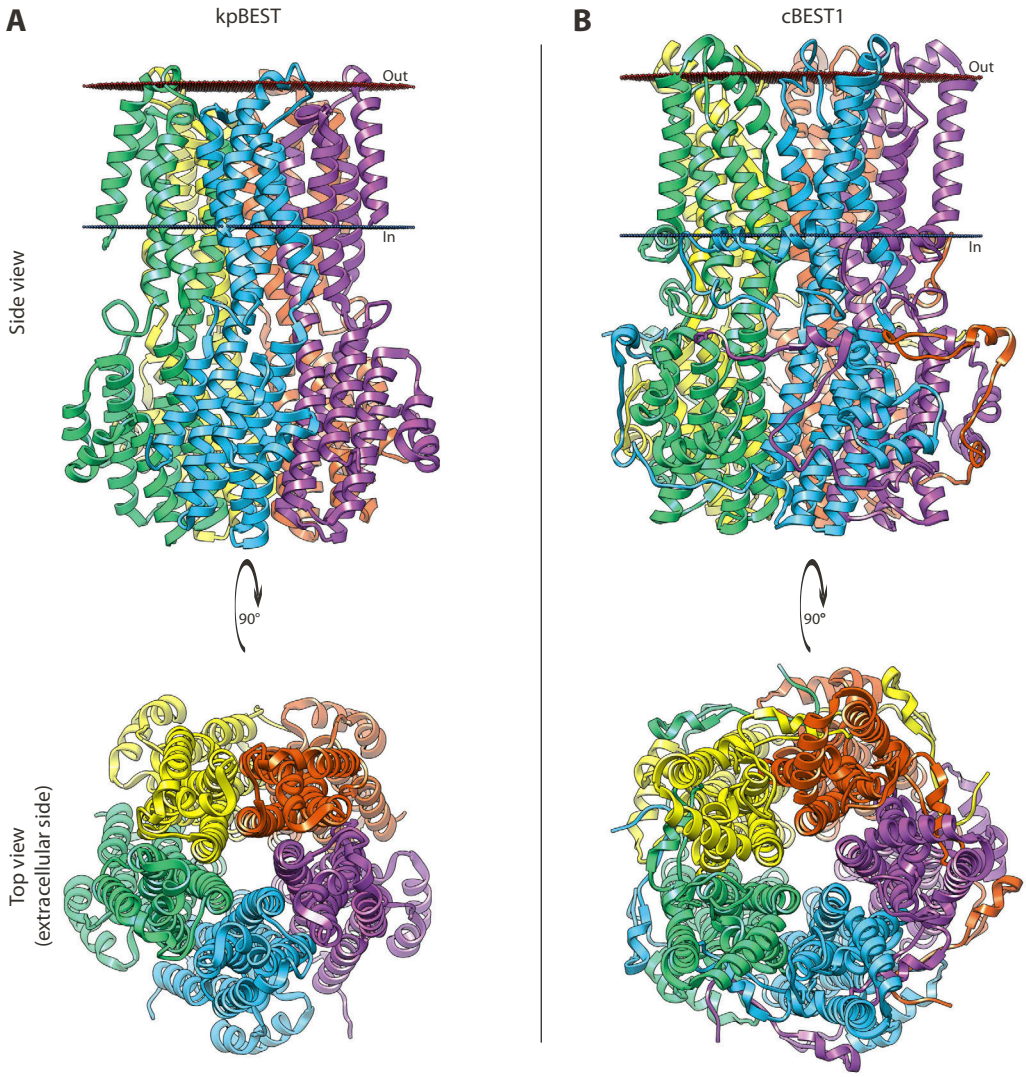


Figure 4 – Ribbon models of kpBEST and cBEST1. The figure shows the pentameric structures of **(A)** kpBEST and **(B)** cBEST1. Within each pentamer, the five monomers are coloured in different colours. The upper part of the figure shows the proteins viewed from the side, with the extracellular sides on top and the intracellular sides on the bottom. Sheets of red and blue balls indicate the predicted boundaries of the hydrocarbon cores of the bilayers. The extracellular sides are labelled "out" and the intracellular sides "in". The lower part of the figure shows the views from the extracellular side (top view), with the pores visible as holes in the middle of the proteins.

1.4.1. Structure of kpBEST

The prokaryotic bestrophin homologue kpBEST was cloned from *Klebsiella pneumoniae* and expressed in *E. coli*. Structures were solved of two C-terminal truncations. A construct where the last 7 amino acid residues had been removed, yielded crystals diffracting to 2.9 Å. Removing the last 11 amino acid residues, resulted in crystals diffracting to 2.3 Å. In the crystal structure, 22–24 residues on the N-terminus (varying between the monomers) and a few residues on the C-terminus could not be resolved.

The structure is a homopentamer (Figure 4a), consisting mainly of α-helices. Four of these helices are membrane spanning. A large cytoplasmic region can be observed and only a small part of the protein is predicted to protrude from extracellular side of the membrane. When viewing the protein from either the extracellular or intracellular sides, a pore can be seen running through the middle of the protein, with the five monomers arranged symmetrically around the pore axis.

1.4.2. Structure of cBEST1

The chicken (*Gallus gallus*) BEST1 homologue cBEST1 is closely related to hBEST1 (Figure 3). The crystallisation construct consisted of the first 405 out of 762 amino acid residues. cBEST1 was expressed in yeast and co-crystallised with a Fab fragment that binds preferentially to the calcium bound form of cBEST1. Crystals were grown in the presence of calcium and diffracted to a resolution of 2.85 Å. In the solved structure residues 2–367 could be modelled. The structure of cBEST1 (Figure 4b) is overall similar to that of kpBEST, since both are homopentamers with a very similar fold. In the membrane region, the structure of cBEST1 appears wider as some of the helices are tilted at a different angle. In the cytoplasmic region some differences are observed as well, since cBEST1 has both a structured N-terminal region before the first transmembrane helix and a longer structured C-terminus that wraps around the cytoplasmic region.

1.4.3. Ion pores

As the architectures of kpBEST and cBEST1 appear very similar, general features of the bestrophin ion pore will be described below, based on the observations from both crystal structures.

In each bestrophin structure, the pore runs in the centre of the protein along the non-crystallographic symmetry axis. The pore is thus perpendicular to the membrane and connects the extracellular side to the intracellular side, as it is expected for an ion channel. Constrictions are found along the pore axis in two separate locations (Figure 5). Three residues on the second transmembrane helix in the so-called “neck” form a long constriction, which is located towards the cytoplasmic side of the membrane. Another, shorter constriction is seen on the exit from the cytoplasmic region of the protein. This constriction is referred to as the “aperture”. The residues forming these constrictions are all hydrophobic, but vary between the two structures, except for the innermost residue in the neck-region, which is a phenylalanine in both structures. Between the two constrictions a large inner cavity is found.

Based on mutational studies of cBEST1 [38], it seems that the residues in the neck-region might be involved in gating of the channel. Mutating the three residues in the neck to alanines, resulted in a structure that overall appeared very similar to the wild-type structure, except for the three mutated residues. This triple-mutant was active in the absence of calcium, indicating

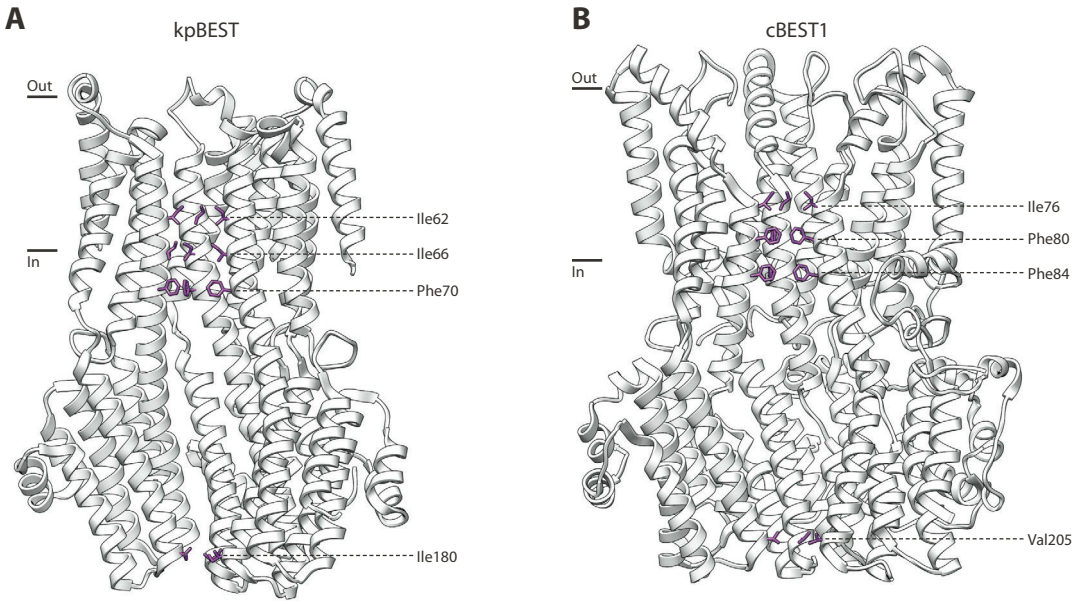


Figure 5 – Views along the pores of kpBEST and cBEST1. The structures of **(A)** kpBEST and **(B)** cBEST1 are shown as white ribbons. From each structure, approximately two whole monomers are hidden, resulting in free views to the pores. Approximate membrane boundaries are shown. The residues causing constrictions along the pores are shown as purple sticks and labelled.

that the neck region might form a gate than can open upon calcium binding. The triple-mutation did not alter the selectivity of the channel, which indicated that the residues in the neck did not form a selectivity filter. Mutating the valine in the aperture to alanine did not change the selectivity for anions over cations, but it did change the size-selectivity. Whereas the relative permeability for different anions in the wild-type channel follows the sequence $SCN^- > I^- > Br^- > Cl^-$, the V205A mutant showed indistinguishable permeabilities for those anions, suggesting that the aperture might form a size-selective filter. It is still not clear what causes cBEST1 to be anion selective, while kpBEST has been reported to be cation selective [67], but one possible explanation could be the presence of anion binding sites that were observed along the pore in the crystal structure of cBEST1 [38].

1.4.4. Ligand binding

kpBEST is reported to have basal activity in the absence of a ligand [67], but it is unknown if the protein can be activated further by some unknown ligand. cBEST1, however, is activated by calcium and, as mentioned previously, calcium was present during the crystallisation. One calcium ion could be observed for each monomer in the structure. The binding site was found to be located after the last transmembrane helix, close to the membrane interface, in a region containing the five highly conserved acidic residues that were previously described. Only two of the conserved acidic residues (Asp301 and Asp304) were involved directly in the binding, while the rest of the binding interactions were contributed by main-chain carbonyls from the surrounding region and the N-terminus and by a water molecule (Figure 6). As expected, mutating the two aspartates directly involved in the calcium binding to alanines affected the ac-

tivation by calcium. Whereas the wild-type channel was activated by Ca^{2+} with an EC50 value of 17 nM, the D301A-D304A mutant could not be activated by 2 µM Ca^{2+} [38].

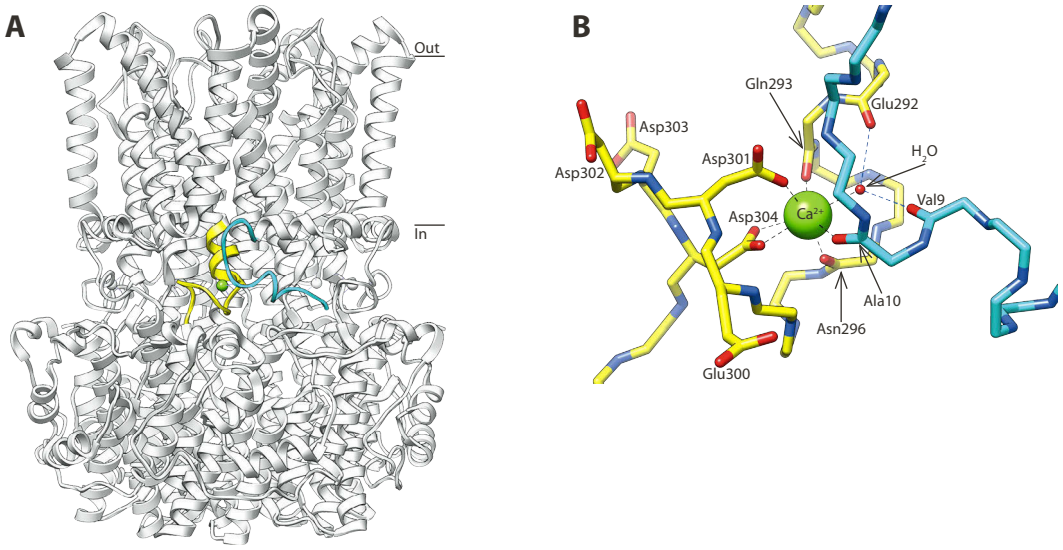


Figure 6 – Ca^{2+} binding site of cBEST1. **(A)** Location of the Ca^{2+} binding site in the structure of cBEST1. N-terminal region coloured cyan, Ca^{2+} binding loop coloured yellow. **(B)** Ca^{2+} binding site of cBEST1. All five acidic residues and the main-chain carbonyls that are involved in Ca^{2+} binding are shown and labelled.

1.5. Aim of this thesis

At the time when the project described in this thesis was started, no structures of bestrophin homologues had been solved. It was also not known if bestrophins were truly channels or some sort of channel regulators. It was therefore my goal to identify a bestrophin homologue that was stable when overexpressed and purified. Using such a homologue, the next goal was to solve a bestrophin structure at atomic resolution and use the structural knowledge to understand the function of bestrophins.

This thesis describes the identification of several prokaryotic bestrophin homologues suitable for structural studies. Against one of these homologues, nanobodies were selected to aid in crystallisation and thus the crystal structure of a prokaryotic bestrophin homologue in complex with a nanobody could be solved. The structure revealed a pentameric ion channel where the ion conduction pathway could readily be identified. Based on the structure, features believed to be important for selectivity and activation could be identified and discussed.

2. Results

2.1. Small scale expression screening

2.1.1. Selection and cloning of prokaryotic homologues

Although bestrophin homologues at first were identified exclusively in eukaryotic organisms, one paper has suggested that prokaryotic homologues also exist. Hagen and colleagues were able to identify 30 putative prokaryotic bestrophin homologues by using the hBEST1 protein sequence as a query for a BLAST search [29]. Based on multiple sequence alignments, these prokaryotic homologues were found to share four short sequence motifs with the eukaryotic homologues. Furthermore, a similar number of transmembrane helices were predicted for prokaryotic and eukaryotic homologues. The prokaryotic homologues were also found to be shorter than the eukaryotic bestrophin homologues, making the prokaryotic homologues interesting targets for structural studies. Eukaryotic membrane proteins are often difficult to purify and crystallise and it was therefore hoped that the more compact prokaryotic homologues would prove to be more amenable to overexpression, purification and crystallisation.

Starting with a collection of prokaryotic genomic DNA, which was acquired for other projects, bacterial bestrophin homologues were identified by searching the NCBI database [68], while limiting the results to organisms that were represented in the available genomic DNA collection. Searching for the keyword "bestrophin", 30 sequences were immediately identified (Table 1). These were sequences that had been automatically annotated as belonging to the bestrophin protein family. BLAST searches did not reveal more homologues within the available collection.

The 30 identified sequences belonged to a total of 18 different organisms. In most organisms, only one homologue was identified, but some contained between one and five additional homologues. To avoid confusion, each homologue was given a unique name, which consisted of a sequentially increasing number and a three-letter abbreviation of the species name (Table 1). All homologues are referred to exclusively by this name in the following. 26 of the 30 homologues were successfully cloned and sequenced. Screening of these homologues is described in a later section.

Since the initial screening was based exclusively on genomic DNA that had been acquired by our lab in order to clone different membrane transporters for other projects, the first round might have been biased towards prokaryotic organisms containing those transporters. It therefore became a concern that the sequence diversity might have been limited in the first round,

Homologue name	Species	NCBI protein accession no.	Length (aa)	Score					
				1	2	3	4	5	6
1SPH	Sphingomonas sp. SKA58	ZP_01302606.1	294	x					
2GAU	Gemmatimonas aurantiaca T-27	YP_002762845.1	337		x				
3GAU	-	YP_002762641.1	285						x
4DFE	Dyadobacter fermentans DSM 18053	YP_003089799.1	335			x			
5DFE	-	YP_003085689.1	333				x		
6DFE	-	YP_003085569.1	293					x	
7DFE	-	YP_003085312.1	297				x		
8DFE	-	YP_003084712.1	302			x			
9DFE	-	YP_003084476.1	300				x		
10PHE	Pedobacter heparinus DSM 2366	YP_003092309.1	286				x		
11CPI	Chitinophaga pinensis DSM 2588	YP_003124670.1	305				x		
12CPI	-	YP_003121699.1	289						x
13CPI	-	YP_003120949.1	305			x			
14SLI	Spirosoma linguale DSM 74	YP_003389066.1	314				x		
15SLI	-	YP_003386700.1	308				x		
16SLI	-	YP_003386305.1	309			x			
17BPS	Bacillus pseudomycoides DSM 12442	ZP_04152149.1	310			x			
18ABO	Alcanivorax borkumensis Yakimov et al. 1998	YP_692019.1	315				x		
19STY	Salmonella typhimurium LT2	NP_460487.1	315		x				
20ECO	Escherichia coli MC1061	YP_001730508.1	304				x		
21MMA	Microscilla marina ATCC 23134	ZP_01689721.1	312	x					
22MMA	-	ZP_01689715.1	326				x		
23MMA	-	ZP_01687254.1	293			x			
24RPA	Rhodopseudomonas palustris TIE-1	YP_001992432.1	307				x		
25MEX	Methylobacterium extorquens PA1	YP_001637861.1	306	x					
26ZPR	Zunongwangia profunda SM-A87	YP_003586571.1	296		x				
27CAL	Cellulophaga algicola DSM 14237	YP_004163015.1	332				x		
28MTR	Marivirga tractuosa DSM 4126	YP_004052572.1	290				x		
29MAN	Maribacter sp. HTCC2170 ('antarcticus')	YP_003862288.1	334	x					
30RBI	Robiginitalea biformata HTCC2501	YP_003193808.1	345				x		

Table 1 – Overview of the first round of screening. Prokaryotic bestrophin homologues selected for screening and results of the screening are listed. For 20ECO, the listed NCBI accession number corresponds to a different strain. Score 1: Cloning failed; score 2: low whole-cell GFP fluorescence; score 3: poorly extracted with DDM; score 4: degraded / not well folded / no FSEC peak; score 5: FSEC peak, but low expression; score 6: FSEC peak, good expression.

as bestrophins could potentially be present in other phylogenetic groups than the transporters selected for other, unrelated projects.

Searching the NCBI database for prokaryotic protein sequences belonging to the bestrophin family gave a list of approximately 700 sequences. From those sequences, a phylogenetic tree was constructed. Homologues from the first round of screening were included in the tree. Homologues were then selected throughout the tree, primarily from organisms where genomic DNA was commercially available. Like in the first round, when a selected organism contained more than one homologue, all of them were chosen for cloning. Additionally, some homologues were chosen based on similarity to the ones that performed well in the first round. Two homologues (69MCA and 70MCA) were found in the thermotolerant prokaryote *Methylococcus capsulatus* (strain Bath), which grows well at temperatures up to 50°C, with the optimum at 37°C [69]. Another homologue (94TEL) was found in the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1, which grows well at temperatures up to 60°C, with the optimum at 57°C [70]. In total, 64 new homologues were selected from 38 different organisms (Table 2). Of these, 56 were successfully cloned and sequenced. Screening of these homologues is described in a later section.

Homologue name	Species	NCBI protein accession no.	Length (aa)	Score					
				1	2	3	4	5	6
31SSP	Sphingobacterium spiritivorum ATCC 33300	ZP_03967349.1	342					x	
32SSP	-	ZP_03966718.1	295			x			
33SSP	-	ZP_03967326.1	294						x
34HHY	Haliscomenobacter hydrossis DSM 1100	YP_004450215.1	328			x			
35HHY	-	YP_004448050.1	292				x		
36HHY	-	YP_004447651.1	308			x			
37RSL	Runella slithyformis DSM 19594	YP_004653851.1	311					x	
38RSL	-	YP_004655318.1	287					x	
39FJO	Flavobacterium johnsoniae UW101	YP_001195730.1	315			x			
40FJO	-	YP_001195282.1	287						x
41FJO	-	YP_001196255.1	305				x		
42BMY	Bacillus mycoides DSM 2048	ZP_04169863.1	310			x			
43PLI	Planctomyces limnophilus DSM 3776	YP_003630025.1	306					x	
44PLI	-	YP_003630739.1	303			x			
45MFU	Myxococcus fulvus HW-1	YP_004668339.1	323	x					
46MFU	-	YP_004665463.1	387	x					
47MFU	-	YP_004665484.1	306	x					
48PMA	Planctomyces maris DSM 8797	ZP_01853801.1	305		x				
49FTA	Fluviicola taffensis DSM 16823	YP_004344559.1	290						x
50MMO	Methylotenera mobilis JLW8	YP_003047530.1	287					x	
51MTU	Methylobacter tundripaludum SV96	ZP_08783233.1	319	x					
52PBR	Planctomyces brasiliensis DSM 5305	YP_004269127.1	302					x	
53BMA	Blastopirellula marina DSM 3645	ZP_01089321.1	295					x	
54BMA	-	ZP_01089414.1	316		x				
55PSY	Pseudomonas syringae pv. aptata str. DSM 50252	EGH76367.1	295						x
56PSY	-	EGH75905.1	314		x				
57XGA	Xanthomonas gardneri (ex Sutic 1957) Jones et al. 2006	ZP_08183953.1	298			x			
58GDI	Gluconacetobacter diazotrophicus PAl 5	YP_001602096.1	291					x	
59CTE	Comamonas testosteroni KF-1	ZP_03544906.1	315				x		
60SMA	Stenotrophomonas maltophilia K279a	YP_001974135.1	305			x			
61OIN	Oceanibulbus indolifex HEL-45	ZP_02155189.1	301				x		
62ABA	Acinetobacter baumannii ATCC 19606	ZP_05829321.1	303					x	
63BTH	Burkholderia thailandensis E264	YP_439619.1	302				x		
64BPH	Burkholderia phymatum STM815	YP_001858405.1	306					x	
65PMI	Proteus mirabilis ATCC 29906	ZP_03841694.1	305			x			
66CTU	Cronobacter turicensis z3032	YP_003210555.1	303				x		
67KPN	Klebsiella pneumoniae subsp. pneumoniae ATCC 13882	ZP_06017540.1	303						x
68MCH	Methylobacterium chloromethanicum CM4	YP_002419264.1	306	x					
69MCA	Methylococcus capsulatus str. Bath	YP_115444.1	277						x
70MCA	-	YP_115096.1	315				x		
71S21	Sphingobacterium sp. 21	YP_004316568.1	293	x					
72S21	-	YP_004319178.1	308	x					
73S21	-	YP_004318133.1	296	x					
74CHU	Cytophaga hutchinsonii ATCC 33406	YP_677610.1	290				x		
75CHU	-	YP_678981.1	297				x		
76CHU	-	YP_678426.1	313				x		
77CHU	-	YP_677323.1	335			x			
78LBY	Leadbetterella byssophila DSM 17132	YP_003996785.1	292					x	
79LBY	-	YP_003997008.1	324			x			
80CGL	Chryseobacterium gleum ATCC 35910	ZP_07089386.1	305				x		
81CGL	-	ZP_07089184.1	292				x		
82CGL	-	ZP_07088791.1	342		x				
83CGL	-	ZP_07086903.1	334			x			
84CGL	-	ZP_07086450.1	292				x		
85CGL	-	ZP_07084811.1	293			x			
86CGL	-	ZP_07084468.1	292		x				
87FPS	Flavobacterium psychrophilum JIP02/86	YP_001295539.1	287				x		
88PSA	Pedobacter saltans DSM 12145	YP_004273412.1	285					x	
89CAL	Cellulophaga algicola DSM 14237	YP_004163015.1	287					x	
90CME	Cupriavidus metallidurans CH34	YP_587474.1	303				x		
91CNE	Cupriavidus necator N-1	YP_004682225.1	338			x			
92CNE	-	YP_004681119.1	285				x		
93RSO	Ralstonia solanacearum GMI1000	NP_521533	306		x				
94TEL	Thermosynechococcus elongatus BP-1	NP_682999	326				x		

Table 2 – Overview of the second round of screening. Prokaryotic bestrophin homologues selected for screening and results of the screening are listed. For 67KPN, the listed NCBI accession number corresponds to a different strain. 89CAL is identical to 27CAL from the first round, except that a later start codon was chosen. 66CTU could not be analysed by FSEC due to technical issues. Score 1: Cloning failed; score 2: low whole-cell GFP fluorescence; score 3: poorly extracted with DDM; score 4: degraded / not well folded / no FSEC peak; score 5: FSEC peak, but low expression; score 6: FSEC peak, good expression.

In the third, and last, round of screening it was attempted to find homologues that were similar to the ones that behaved well in the second round. Approximately 1500 sequences were downloaded from the NCBI database and used to construct a phylogenetic tree. Within this tree, a clade was identified which contained several of the homologues that had behaved well in the previous rounds (3GAU, 12CPI, 33SSP, 40FJO and 69MCA). In total, 63 homologues were present in this clade, of which 25 had been tested in the previous rounds. A few additional sequences were found by BLAST searches using 33SSP as a query sequence. This way a total of 34 new homologues were selected for screening (Table 3). Unlike the previous rounds, other homologues from the selected organisms were omitted if they were not belonging to this clade. 13 of the selected homologues could be cloned from commercially available genomic DNA. For the remaining 21 homologues, the corresponding genes were synthesised. All 34

Homologue name	Species	NCBI protein accession no.	Length (aa)	Score					
				1	2	3	4	5	6
95FIN	Flavobacterium indicum GPTSA100-9	YP_005357091.1	287						x
96NKO	Niastella koreensis GR20-10	YP_005006142.1	293					x	
97PAG	Pedobacter agri PB92	ZP_11219306.1	286					x	
98FAE	Fibrella aestuarina BUZ 2	YP_007323946.1	292					x	
99FAE	-	YP_007320478.1	311			x			
100FLI	Fibrisoma limi BUZ 3	ZP_10333830.1	304					x	
101FLI	-	ZP_10333173.1	309			x			
102EOL	Emticicia oligotrophica DSM 17448	YP_006871835.1	290					x	
103OHO	Owenweeksia hongkongensis DSM 17368	YP_004990241.1	289						x
104FDU	Fluoribacter dumoffii NY23	ZP_10139395.1	284				x		
105SGR	Saprospira grandis DSM 2844	ZP_18267275.1	309		x				
106GOB	Gemmata obscuriglobus UQM 2246	ZP_02732374.1	281		x				
107CAN	Cesiribacter andamanensis AMV16	ZP_23804627.1	300					x	
108FBR	Flavobacterium branchiophilum FL-15	YP_004844850.1	287		x				
109FCO	Flavobacterium columnare ATCC 49512	YP_004941078.1	287					x	
110FFR	Flavobacterium frigoris PS1	ZP_09896637.1	285		x				
111LLO	Legionella longbeachae NSW150	YP_003455895.1	287				x		
112LPN	Legionella pneumophila subsp. pneumophila ATCC 43290	YP_005184430.1	286		x				
113FBA	Flavobacteria bacterium BAL38	ZP_01735098.1	287						x
114MSA	Mariniradius saccharolyticus AK6	EMS34534.1	293		x				
115FSP	Flavobacterium sp. CF136	ZP_10730490.1	287					x	
116CSP	Chryseobacterium sp. CF314	ZP_10726656.1	293				x		
117ASP	Algoriphagus sp. PR1	ZP_07722495.1	293				x		
118EAN	Elizabethkingia anophelis Ag1	ZP_09414660.1	290				x		
119FSP	Flavobacterium sp. F52	ZP_10480769.1	287						x
120PSP	Pedobacter sp. BAL39	ZP_01882510.1	286					x	
121LDR	Legionella drancourtii LLAP12	ZP_09618728.1	284				x		
122FSP	Flavobacterium sp. WG21	AMYW01000055	287					x	
123PCU	Flavobacterium sp. Leaf359	AHJF01008472	288					x	
124FSP	Flavobacterium sp. B17	BACY01000542	293				x		
125LAN	Legionella anisa str. Linanisette	CANP01000031	282				x		
126PSP	Pontibacter sp. BAB1700	ZP_10400757.1	286					x	
127FIM	Fulvivirga imtechensis AK7	ZP_20982534.1	290				x		
128FBA	uncultured Flavobacteriia bacterium	CCF99166	290					x	

Table 3 – Overview of the third round of screening. Prokaryotic bestrophin homologues selected for screening and results of the screening are listed. For the homologues 122FSP, 123PCU, 124FSP and 125LAN protein sequences were not available in the NCBI database, when the list was compiled. For those homologues, the corresponding DNA sequences are listed instead. The DNA sequence for the homologue 123PCU is a whole genome shotgun sequence from the organism Pseudoperonospora cubensis, but the cloned sequence actually belongs to the organism listed in the table. For 110FFR, 127FIM and 128FBA an earlier start codon was chosen than the one annotated in the NCBI database. For 104FDU, the listed NCBI accession number corresponds to a different strain. Score 1: Cloning failed; score 2: low whole-cell GFP fluorescence; score 3: poorly extracted with DDM; score 4: degraded / not well folded / no FSEC peak; score 5: FSEC peak, but low expression; score 6: FSEC peak, good expression.

homologues were successfully cloned and sequenced. Screening of these homologues is described in a later section.

2.1.2. Screening methodology and workflow

Green fluorescent protein (GFP) fusion proteins have previously been shown to be useful for expression and purification screening of integral membrane proteins. With GFP fusions, small amounts of protein can be detected in a size-exclusion chromatography (SEC) experiment, by using a liquid chromatography system coupled to a fluorescence detector. This technique, known as fluorescent-detection size-exclusion chromatography (FSEC), makes it possible to evaluate the monodispersity and expression level of small amounts of protein in crude detergent extracts [71]. Monodispersity, defined as a single, symmetrical peak in a size-exclusion chromatography experiment, is generally considered to be an important factor for protein crystallisation [72].

GFP can additionally also be used as a folding indicator. When fused to the C-terminus of a target protein, the correct folding of GFP is typically dependent on the correct folding of the target protein [73]. Folded protein and aggregated protein can then be separated by SDS-PAGE, where the aggregated protein tend to migrate with a higher apparent molecular weight than the correctly folded protein [74].

Since all the selected homologues were from prokaryotic organisms, *Escherichia coli* was chosen as the host for the expression screening. As an expression vector the plasmid pBXC3GH was used. This vector encodes C-terminal GFP and 10×His fusion tags as well as a HRV 3C cleavage site, allowing removal of those tags by proteolytic cleavage with the HRV 3C protease. In this vector, expression of the fusion protein is under the control of the L-arabinose inducible P_{araBAD} promoter. Since expression under the control of this promoter is known to be tightly controlled and can be modulated over a wide range of inducer concentrations [75], it was thought to be ideal for the expression of integral membrane proteins, which are often difficult to produce.

While the first round of screening only included 30 homologues, it was considered likely that more homologues would be screened at a later point. For that reason, a high-throughput expression and screening method was chosen. Different cultures were grown in parallel in small volumes (typically 1–4 ml in 96 or 24-well plates) with subsequent cell lysis, detergent extraction and various centrifugation steps also carried out in parallel.

E. coli cells were generally induced at 25°C as this temperature had been found optimal for other projects [76,77]. In the initial round of screening, several different concentrations of inducer were used. As the concentration of 0.004% L-arabinose was found to work well, this was used for subsequent rounds. After overnight incubation, the cells were harvested by cen-

trifugation. Whole cell GFP fluorescence was used to identify poorly expressing homologues, which were omitted from subsequent steps.

After lysing the cells, the detergent n-Dodecyl-β-D-maltoside (DDM) was added to extract the membrane proteins. Non-solubilised material was removed by ultracentrifugation and the extraction efficiency was judged by comparing the GFP fluorescence of the lysate before (total protein fraction) and after (soluble protein fraction) the ultracentrifugation. Homologues with poor extraction efficiency (below 30–40%) were excluded at this point.

Samples of the total and soluble protein fractions were subsequently analysed by SDS-PAGE. Using in-gel fluorescence, bands of correctly folded fusion protein could be identified. After recording the in-gel fluorescence, the same gels were analysed by western blotting, which made it possible to visualise the misfolded protein in addition to the bands of correctly folded protein. Since all homologues were similar in length (average 303 amino acid residues; SD = 17 amino acid residues), the expected molecular weight of a GFP fusion was around 60 kDa. This knowledge was used to distinguish full-length fusion proteins from degradation products when analysing the gel images. Homologues that were predominantly misfolded or degraded were thus excluded at this step.

Finally, samples of the soluble protein fractions were analysed by FSEC. Most of the tested homologues gave no clear peak, but certain homologues resulted in a single, monodisperse peak that eluted at the expected volume. Homologues that resulted in very small peaks, barely distinguishable from the background signal, were generally not investigated further.

2.1.3. First round of screening

The aim of the first screening round was primarily to test whether bacterial bestrophin homologues could be expressed in *E. coli* and purified, since no published study at that time had investigated this question. Initial results were positive, since 23 of the 26 tested homologues resulted in green fluorescent *E. coli* cells. Of the remaining 23 homologues, 17 could be extracted efficiently with DDM. Some produced mainly aggregated protein, which was seen as a band of higher molecular weight in the western blot analysis. Others seemed to produce well-folded protein, but did not give a single peak in the FSEC analysis. Results for the different homologues are summarised with a score in table 1.

Two homologues behaved very well, each showing a single peak on the FSEC chromatogram. The results for these two homologues, 3GAU and 12CPI, are shown in figure 7. The homologue 20ECO (from *E. coli*) is included as an example of a poorly behaving homologue. As both 3GAU and 12CPI were fused to GFP, the calculated molecular weights were 61 kDa and 63 kDa, respectively. The positive control, LacS, by comparison had a slightly higher calculated molecular weight of 83 kDa with GFP fused to the C-terminus and 54 kDa when only fused

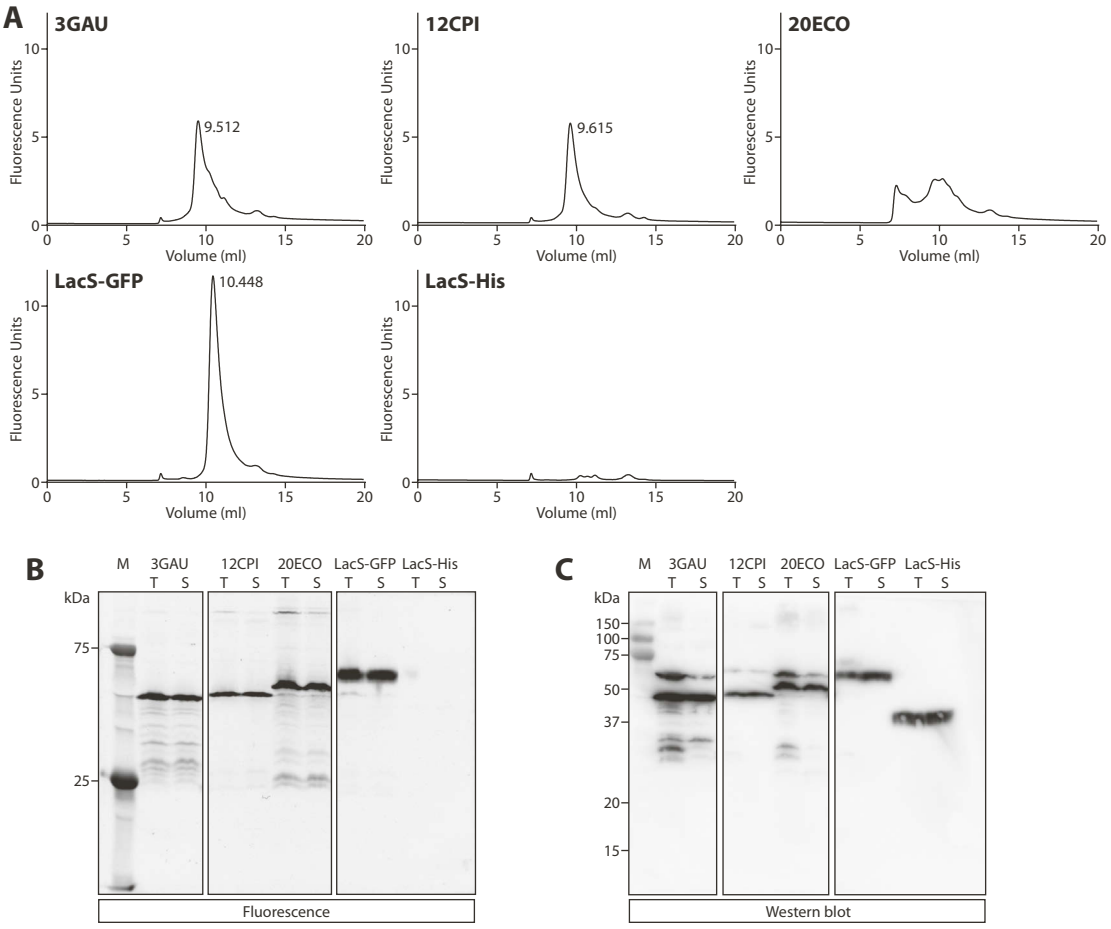


Figure 7 – Results of the first round of small scale expression screening. **(A)** FSEC traces for three of the tested homologues. 3GAU and 12CPI were selected for further testing, whereas 20ECO was not. LacS-GFP and LacS-His were included as positive and negative controls, respectively. Numbers next to the peaks indicate the elution volumes of the peaks. **(B)** In-gel fluorescence results for the same three homologues. The Lane "M" indicates a prestained molecular weight marker. Lanes labelled "T" show the total protein content in the lysate before ultracentrifugation and "S" show the soluble protein left in the supernatant after ultracentrifugation. **(C)** Western blot analysis of the same gel as shown in the previous panel. Detection with anti-His-HRP antibody. The bands of the prestained marker are not all visible on the shown images, but they were visible on the membrane and their positions are shown.

to a C-terminal His-tag. 3GAU and 12CPI, however, eluted earlier than LacS-GFP, which indicated that these proteins might form higher oligomeric structures. 3GAU did not produce a monodisperse peak, since there was a shoulder on the trailing edge of the peak. By comparison, the homologue 20ECO (the homologue from *E. coli*), did not show a well-defined peak.

As illustrated by the in-gel fluorescence analysis in figure 7, 3GAU and 12CPI both showed a clear fluorescent band at a lower molecular weight than LacS. As it is often seen for integral membrane proteins, all proteins migrated faster than their corresponding sizes of marker proteins. 20ECO (64 kDa) also showed a clear fluorescent band. Both 3GAU and 20ECO showed several lower molecular weight bands, which might indicate partial degradation.

3GAU and 20ECO also both showed significant amounts of misfolded protein, resulting in characteristic double bands on the western blot. The higher molecular weight bands were

thought to contain misfolded GFP and were therefore not seen in the in-gel fluorescence images. Most of the misfolded protein was removed by ultracentrifugation, since the intensities of the higher molecular weight bands were greatly reduced after this purification step.

For the two homologues 3GAU and 12CPI, the in-gel fluorescence and the western blots were in good agreement with the FSEC results. 12CPI produced a much cleaner signal in the in-gel analysis and showed a monodisperse peak in the FSEC analysis. For 3GAU, the presence of lower molecular weight bands observed on the gels could explain the FSEC behaviour and also suggested that the protein was sensitive to degradation and therefore might require the addition of protease inhibitors during purification. The presence of misfolded protein before ultracentrifugation suggested that the inducer concentration might have been too high, since this often leads to misfolded protein [74].

Based on the in-gel analysis alone, it could not have been anticipated that 20ECO would not produce a peak in the FSEC analysis. This underlines the importance of performing the FSEC analysis in addition to the gel-based expression screening.

2.1.4. Second round of screening

In comparison to the first round, the second round of screening included significantly more homologues. Of 56 proteins investigated, 50 showed clear fluorescent signals in intact *E. coli* cells. 36 proteins could be extracted with DDM and were investigated further. 17 proteins did not yield sufficient amounts of well folded protein in the gel analyses or did not show single peaks in the FSEC analysis, and were therefore not studied further. Of the remaining 19 proteins, 13 only resulted in very small FSEC peaks and were therefore not studied further, as higher priority was given to the remaining 6 homologues with higher yields. Table 2 summarises the results for the entire set of tested homologues.

Results for the best behaving homologues are shown in figure 8. The homologue 12CPI from the previous round was included as a control. Since the FSEC analysis in this round of screening was done on a different SEC column, all peaks eluted earlier compared to the first round. 67KPN (from *Klebsiella pneumoniae*) is closely related to the *E. coli* homologue 20ECO, which was tested in the first round of screening. The two proteins share a sequence identity of 56% and cluster in proximity in the phylogenetic tree (Figure 9). However, unlike 20ECO, 67KPN showed a single monodisperse peak. Additionally, the gel-analysis showed a clear band and no significant amounts of misfolded protein or degradation products.

As mentioned previously, the homologue 69MCA was cloned from a thermotolerant organism. The FSEC and the gel analyses indicated that this protein behaved well and expressed in larger amounts than the control.

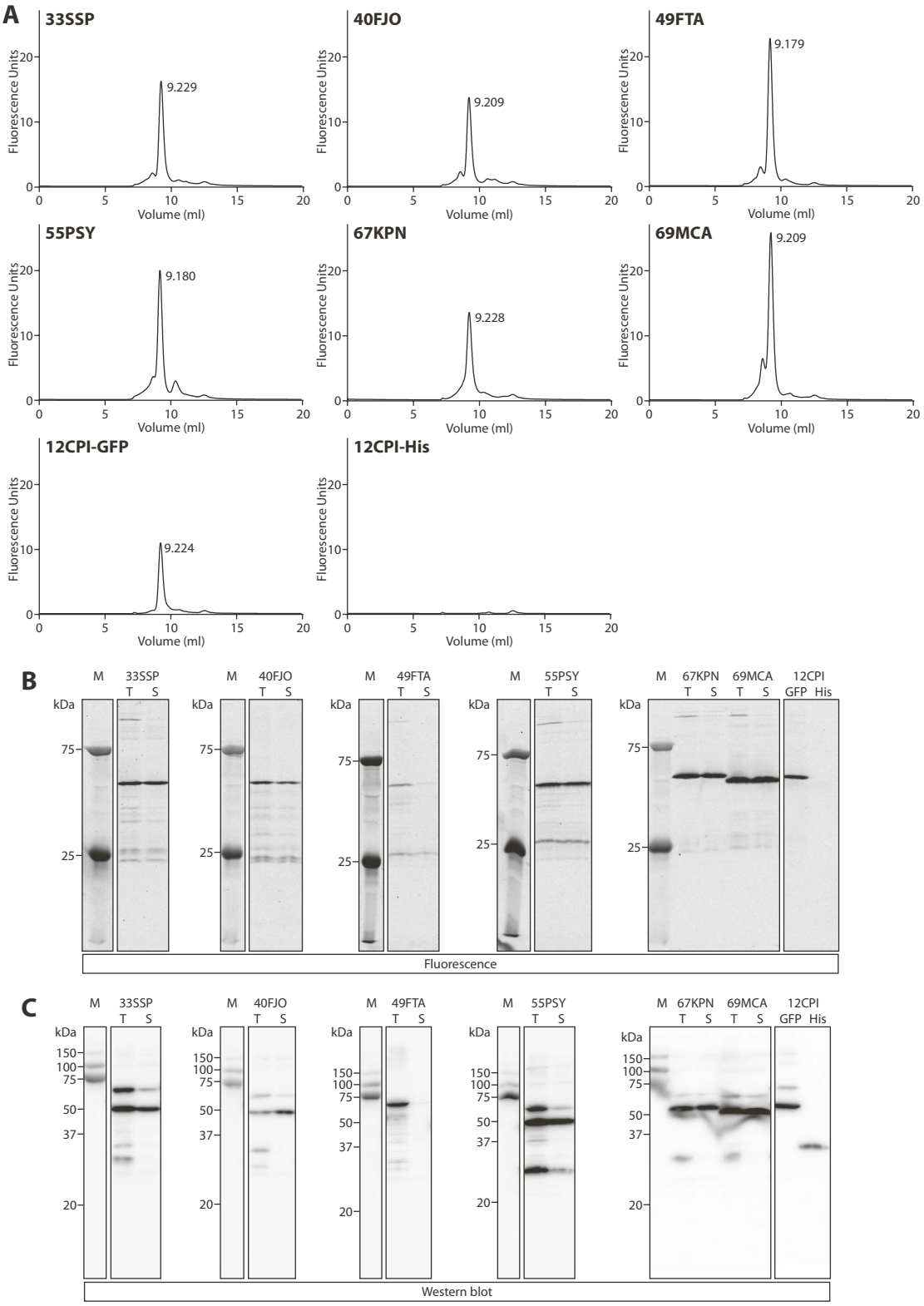


Figure 8 – Results of the second round of small scale expression screening. **(A)** FSEC traces for six well behaving homologues. 12CPI-GFP and 12CPI-His were included as positive and negative controls, respectively. Numbers next to the peaks indicate the elution volumes of the peaks. **(B)** In-gel fluorescence results for the same six homologues. The Lane "M" indicates a prestained molecular weight marker. Lanes labelled "T" show the total protein content in the lysate before ultracentrifugation and "S" show the soluble protein left in the supernatant after ultracentrifugation. **(C)** Western blot analysis of the same gels as shown in previous panel. Detection with anti-His-HRP antibody. The bands of the prestained marker are not all visible on the shown images, but they were visible on the membranes and their positions are shown.

The homologues 33SSP, 40FJO and 55PSY all behaved promising, although 33SSP and 55PSY did show some misfolded protein before the ultracentrifugation step. 55PSY also showed an additional band at a lower molecular weight, indicating degradation. Both homologues, however, gave good results in the FSEC analysis.

49FTA gave a single monodisperse peak in the FSEC analysis, but only a faint band was visible on the gels after ultracentrifugation. It was still included in the further analysis, since the FSEC peak indicated a high expression level.

2.1.5. Third round of screening

34 homologues were tested in the third and final round of screening. Results are summarised in table 3. Judged by the whole cell fluorescence, 6 homologues did not express well. Of the re-

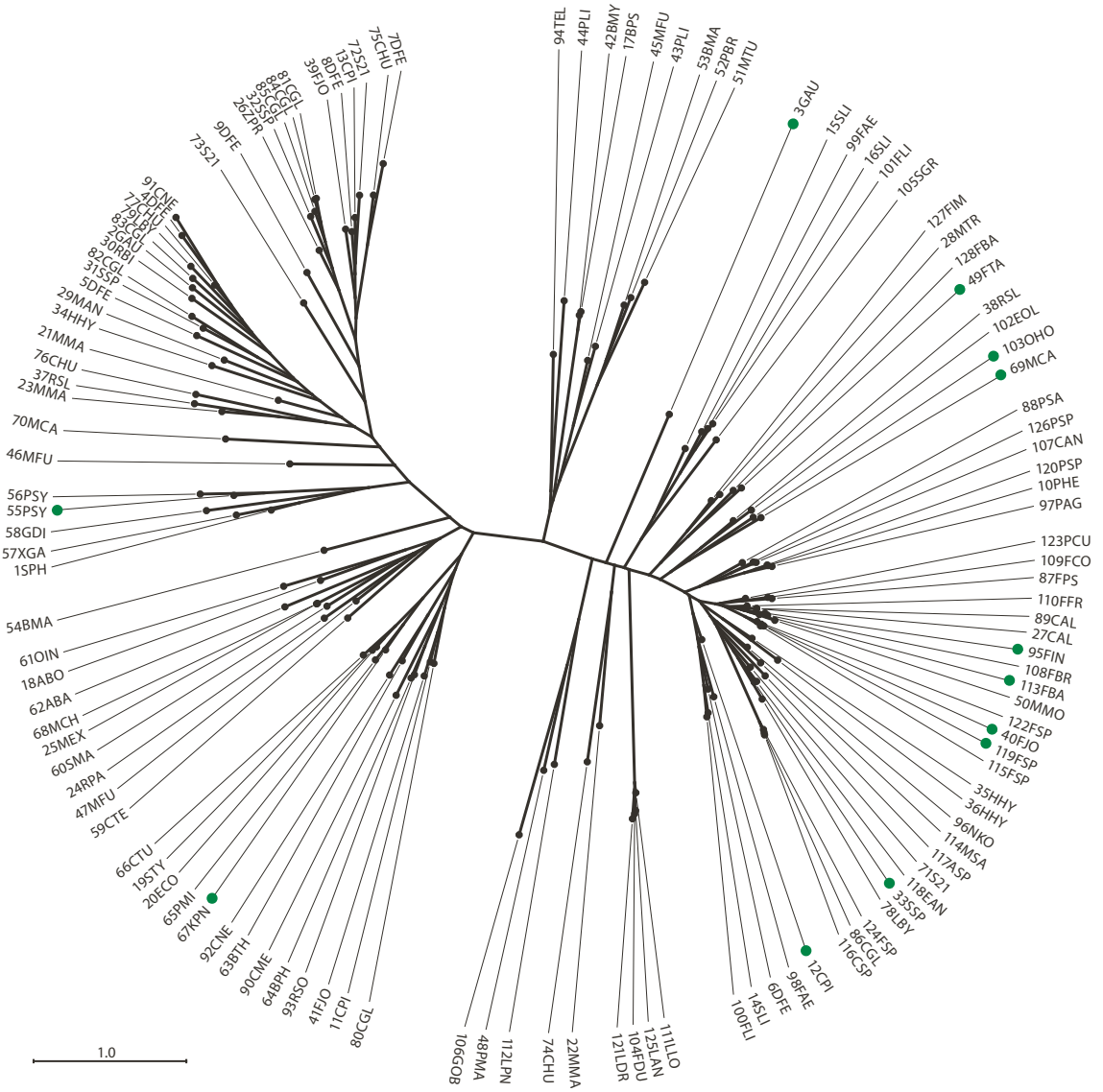


Figure 9 – Maximum likelihood phylogenetic tree showing the phylogeny of all tested homologues. The tree was calculated using PhyML. The scale bar indicates the average number of amino acid substitutions per site. Homologues that performed well in the small scale expression screenings are highlighted with green dots.

maining 28 homologues, 2 could not be extracted and 9 did not show any signal during FSEC analysis.

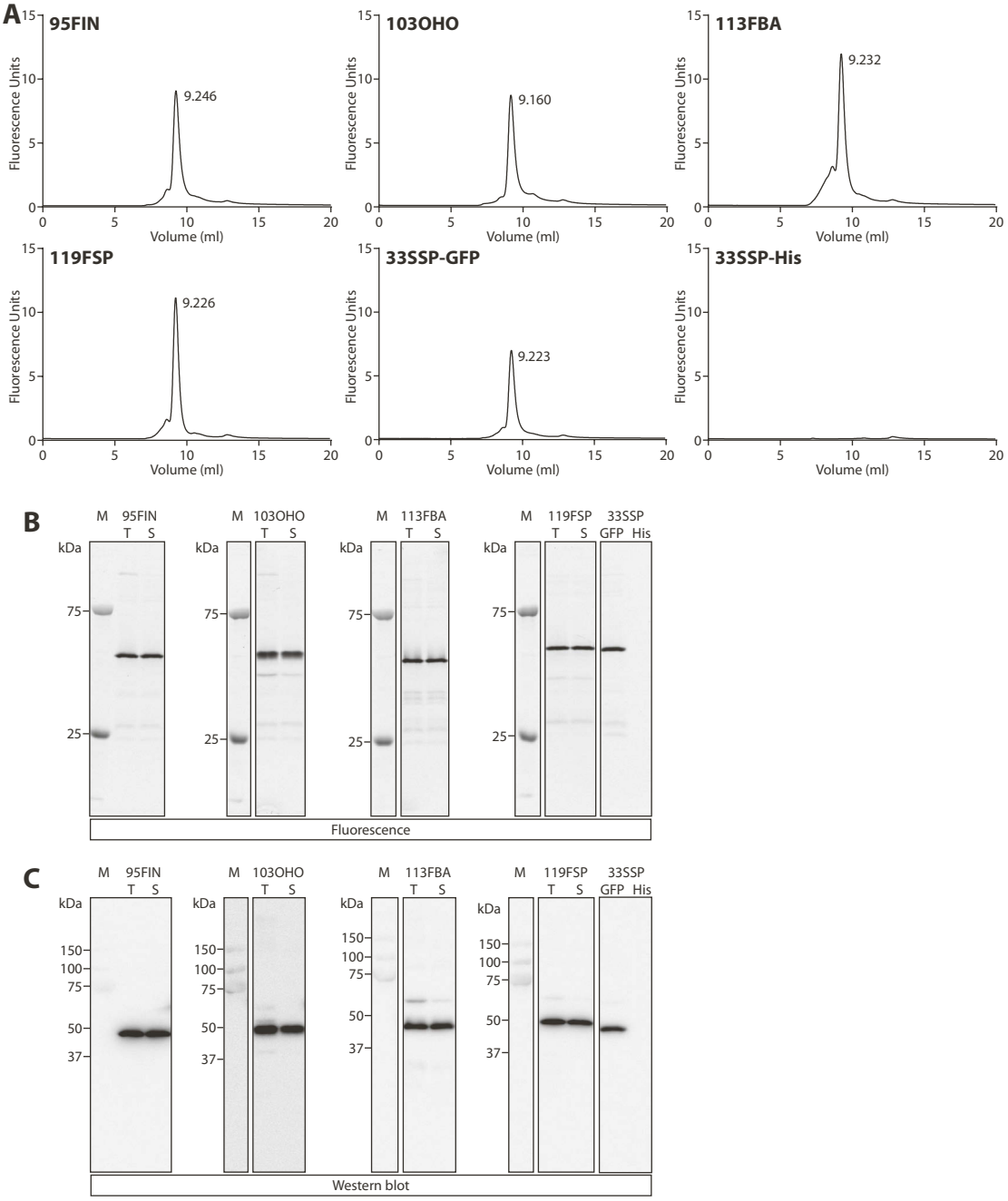


Figure 10 – Results of the third round of small scale expression screening. **(A)** FSEC traces for four well behaving homologues. 33SSP-GFP and 33SSP-His were included as positive and negative controls, respectively. Numbers next to the peaks indicate the elution volumes of the peaks. **(B)** In-gel fluorescence results for the same four homologues. The Lane "M" indicates a prestained molecular weight marker. Lanes labelled "T" show the total protein content in the lysate before ultracentrifugation and "S" show the soluble protein left in the supernatant after ultracentrifugation. **(C)** Western blot analysis of the same gels as shown in previous panel. Detection with anti-His-HRP antibody. The bands of the prestained marker are not all visible on the shown images, but they were visible on the membranes and their positions are shown. Less marker was loaded than in the first two rounds of screening and as a result of that, the 20 kDa band of the marker was not clearly visible. 33SSP did not express well with only a His-tag and therefore no band is visible in the lane labelled 33SSP-His.

From this screening round, in total 17 homologues resulted in FSEC peaks eluting at the expected volume. This was almost as many as the second round of screening, where peaks were observed for 19 homologues. Considering that significantly more homologues were tested in the second round, this meant that the success rate was higher in the third round. This is likely an effect of choosing homologues closely related to the ones that were performing well in previous rounds.

The homologues that expressed the best and gave the cleanest gel signals, without signs of much degradation or misfolding, were selected for further studies. Figure 10 shows the results of FSEC and gel analyses for the homologues 95FIN, 103OHO, 113FBA and 119FSP. Based on the FSEC elution profiles, all four homologues expressed at least at the same level or even higher than the homologue 33SSP, which was included as a control for comparison. Peaks were symmetric and monodisperse.

For each of the four homologues a single band was visible in the in-gel fluorescence analysis. Lower molecular weight bands, indicating degradation, were faint. Similarly, the western blot analysis showed little misfolded protein, which could be removed by ultracentrifugation.

The sequences of the four homologues selected for further characterisation were aligned to the sequences from the previous rounds of screening and are shown in figure 11. The sequences aligned well, except for a few regions where some homologues were significantly shorter than others.

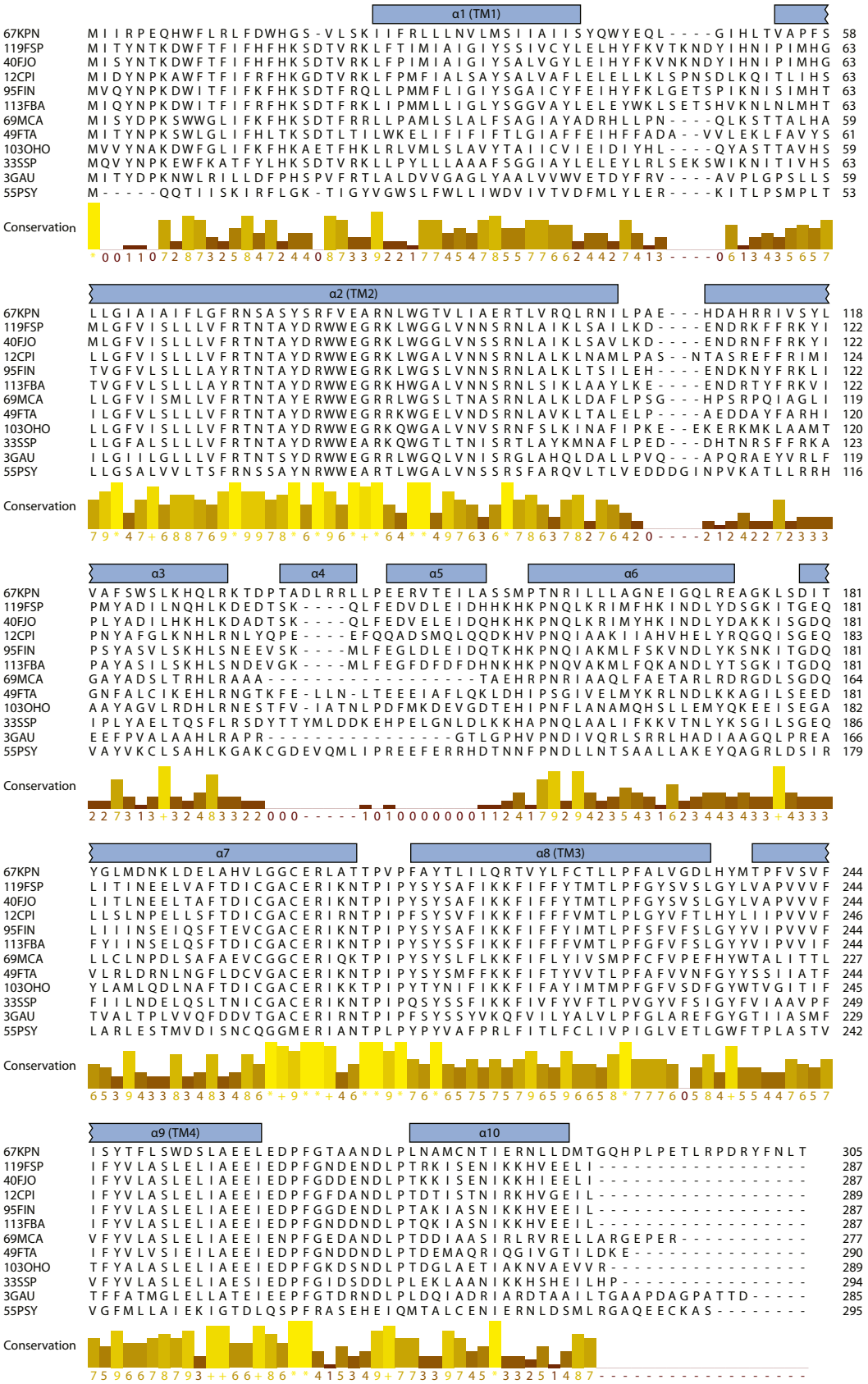


Figure 11 – Multiple sequence alignment of the homologues selected during initial screening. The alignment was prepared with PROMALS3D. α-helices are shown as blue rectangles for the homologue 67KPN based on the PDB entry 4WD8. The membrane spanning helices are labelled as TM1–4.

2.2. Characterisation of selected homologues

2.2.1. Choosing detergents and additives

Integral membrane proteins are typically crystallised as detergent-protein complexes. A detergent molecule consists of a hydrophilic head group and a hydrophobic tail. In aqueous solutions, detergents can spontaneously assemble into micelles when the concentration is sufficiently high. In these micelles, the hydrophobic tails are in the middle and the hydrophilic head groups on the surface. When a membrane protein is solubilised in a detergent, the detergent micelle surrounds the hydrophobic region of the protein like a belt. In most membrane protein crystals, the protein stays inside the detergent micelle [78].

The size of a micelle can vary dependent on the type of the detergent head group as well as the length of the detergent alkyl chain. Shorter chain detergents generally form smaller micelles than longer chain detergents. When crystallising detergent solubilised membrane proteins, it is generally thought that larger micelles can hinder the formation of crystal contacts and therefore result in crystals that diffract less well [79]. Unfortunately, shorter chain detergents are generally harsher than longer chain detergents, so a balance must be found between the improved crystallisation properties of shorter chain detergents and the improved protein stability of the longer chain detergents.

DDM, which was used in the small-scale screening, is a mild detergent which has been widely used for the purification and crystallisation of membrane proteins. The detergent n-Decyl-β-D-maltoside has the same maltoside head group as DDM, but has an alkyl chain that is two carbon atoms shorter. Thus, DM produces smaller micelles than DDM, but it is also harsher. Since bacterial membrane proteins are often more stable than comparable eukaryotic homologues [79], I expected that the shorter chain detergent DM could be a good choice for crystallisation of the bacterial bestrophin homologues 3GAU and 12CPI.

Animal bestrophins are known to form calcium activated ion channels. Nothing was known about the activation of prokaryotic bestrophin homologues, but it was considered a possibility that they could also be activated by calcium. Therefore, crystallisation screens were, whenever possible, performed both with and without the addition of calcium. Typically, one protein drop contained no added calcium and another contained 1–5 mM $CaCl_2$.

2.2.2. Purification and crystallisation of homologues from the first round

For their crystallisation, several milligrams of protein are typically required to screen thousands of different conditions at once. Therefore, the expression of 3GAU and 12CPI had to be scaled up to larger volumes of *E. coli* culture. Essentially the expression was carried out as during the small-scale screening, except that larger culture vessels were used. Typically,

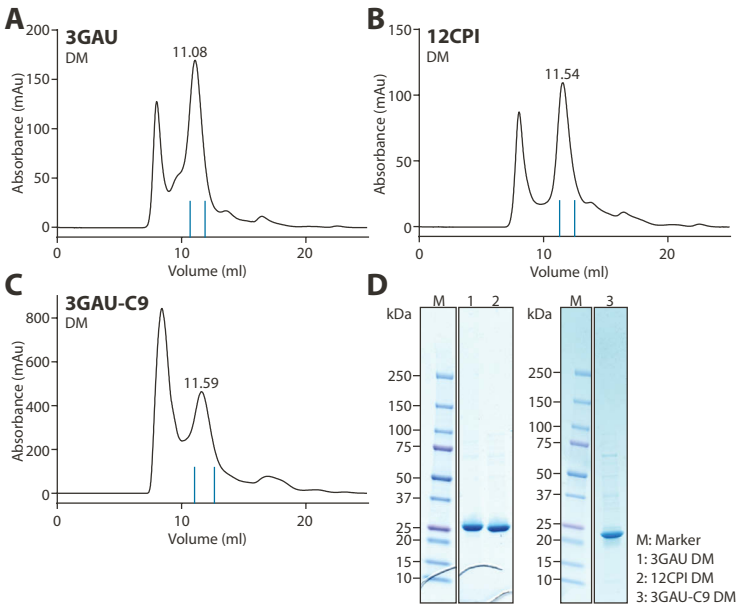


Figure 12 – Purification of 3GAU, 12CPI and 3GAU-C9 in DM. **(A)** 3GAU. **(B)** 12CPI. **(C)** 3GAU-C9. The shown chromatograms are SEC profiles. Numbers on the chromatograms indicate elution volumes of the peaks. Blue lines on the bases of the chromatograms indicate the peak fractions, which were loaded on the shown gels. **(D)** SDS-PAGE analyses of the peak fractions.

cultures were grown in 2 l shaker flasks or in a fermenter. Typical volumes were 32 × 0.6 l in shaker flasks or 9 l in a fermenter.

To purify the protein, a two-step protocol was developed. In the first step the protein was purified using immobilised metal ion affinity chromatography (IMAC), where the C-terminal His-tag of the protein was bound to Ni^{2+}-ions that were immobilised on agarose beads. After proteolytic removal of the GFP–His-tag, the protein was then purified using size-exclusion chromatography (SEC).

Since initial attempts to purify 3GAU and 12CPI using the detergent DDM looked promising (data not shown), I decided to attempt to purify the homologues in DM instead, hoping that the shorter micelle size of DM would be beneficial during crystallisation screening. Figure 12a–b shows the size-exclusion chromatograms from purifications using approximately 1–1.5 l fermenter culture. For both homologues, a comparably large peak was obtained at the void volume, followed by a single, monodisperse peak eluting after 11.08 ml (3GAU) or 11.54 ml (12CPI). When comparing to reference protein standards, these elution volumes would correspond to soluble, globular proteins with sizes ranging between 290 kDa (12CPI) and 360 kDa (3GAU). Since the hydrodynamic radius of a protein-detergent-complex depends both on the amount of protein and the size of the detergent micelle, and since the size of a detergent micelle surrounding a membrane protein is not exactly known, the exact oligomeric state of the proteins could not be determined based solely on the SEC profiles. Nevertheless, the

chromatograms show that the proteins assembled into higher-order oligomers, consisting of several identical chains. As shown in figure 12d, both protein samples were pure.

Several attempts were made to crystallise 3GAU and 12CPI purified in DM, but no crystals were obtained. 3GAU was found to have a C-terminus that was 13 amino acid residues longer than that of 12CPI. A shorter version of the C-terminus of 3GAU, which was constructed by removing the last 9 residues (referred to as 3GAU-C9), was found to be expressed at similar levels as the full-length protein and to be stable during purification. Figure 12c shows the size-exclusion chromatogram from a purification of 3GAU-C9, which was used for crystallisation screening. Despite several attempts, 3GAU-C9 did not crystallise in DM either. Also, the addition of calcium did not seem to have any effect on crystallisation.

Interestingly, when purifying larger amounts of protein, I often observed that the void peak was comparably large compared to the main protein peak (e.g. the purification in figure 12c). This observation led to the question of how stable these homologues were in DM. To investigate this, a simple stability experiment was performed with purified 3GAU-C9 as shown in figure 13. Small amounts of protein from a size-exclusion peak was incubated either at room temperature or at 4°C. After 24 hours at room temperature, a large fraction of the protein had aggregated, indicating a limited stability of the protein in DM. Since there was also aggregation of the protein at 4°C the suitability of this homologue for crystallisation studies in DM was questionable.

With the hope of improving its stability, crystallisation screening was subsequently carried out with 3GAU-C9 and 12CPI purified in DDM. Both crystallised at 4°C in various conditions, predominantly at low concentrations of PEG400. Unfortunately, the crystals diffracted to low resolution. The addition of calcium did not have any effect on crystallisation or diffraction.

Figure 14 shows a diffraction image from the best diffracting crystal of the homologue 3GAU-C9. Diffraction spots were visible to a resolution of about 8 Å. Most other crystals diffracted to lower resolution with less well-defined diffraction patterns. Details of the data

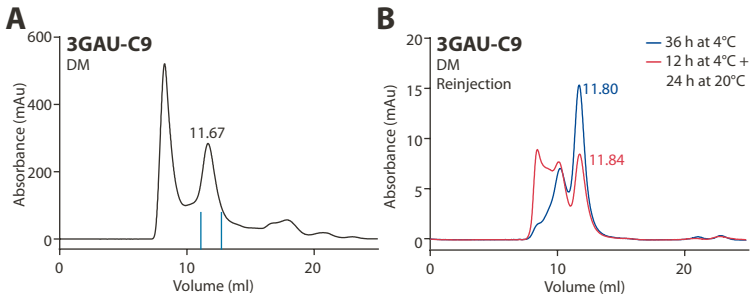


Figure 13 – Stability test of 3GAU-C9 in DM. **(A)** SEC profile from purification of 3GAU-C9 in DM. Blue lines on the base of the chromatogram indicate the peak fraction. A part of the peak fraction was split into separate tubes that were incubated at different temperatures. **(B)** Samples were reanalysed by SEC after 36 h at 4°C (blue curve) or 12 h at 4°C + 24 h at 20°C (red curve).

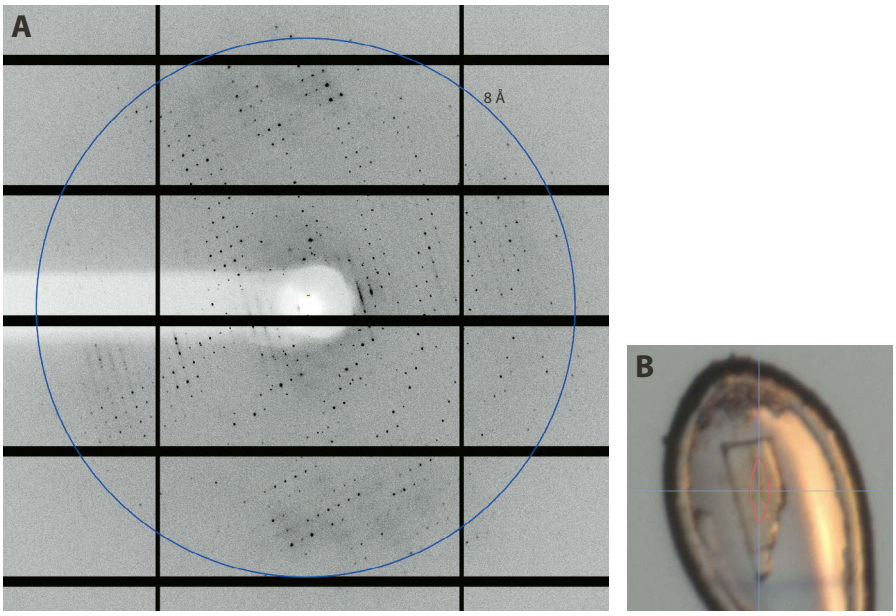


Figure 14 – Best diffracting crystal of 3GAU-C9. **(A)** Observed diffraction pattern. Blue ring indicates the 8 Å resolution ring. **(B)** Picture of the crystal in the loop, mounted at the beamline.

collection are shown in table 4. The data was complete up to a resolution of 7.8 Å. Based on the size of the unit cell, it can be assumed that the crystal contains several copies of a large membrane protein–detergent complex. This experiment showed that the crystallisation of bacterial bestrophins was possible, but the limited diffraction quality of the crystals meant that focus should be shifted towards the homologues identified in the second round of screening, in the hope to identify better behaved homologues.

Crystal	
Protein:	3GAU-C9 in DDM
Additive:	5 mM $CaCl_2$
Condition:	500 mM KCl
	50 mM HEPES pH 7.4
	15% PEG400
Data collection	
Beamline:	X06SA, SLS
Wavelength:	1.000 Å
Space group:	P 1 2_1 1
Cell dimensions:	
a, b, c (Å)	148.23, 253.76,194.28
α, β, γ (°)	90.00, 100.80, 90.00
Resolution (Å):	49.30 – 7.80 (7.93 – 7.80)
R_{merge} (%):	9.2 (93.9)
R_{meas} (%):	10.1 (101.2)
$CC_{1/2}$:	0.997 (0.734)
$\|I\|/\sigma(I)$:	10.5 (2.0)
Completeness (%):	99.76 (99.87)
Redundancy:	6.7

Table 4 – Details of crystallisation and data collection statistics for the best diffracting crystal of 3GAU-C9. Numbers in parentheses indicate the highest resolution shell and its statistics.

2.2.3. Initial purification of homologues from the second round

Based on the experience with the homologues 3GAU and 12CPI, all of the selected homologues from the second round of screening were expressed and purified the same way. As a detergent DDM was chosen, since it had resulted in the best crystals of 3GAU. The size-exclusion chromatograms from the purifications together with images of the peak fractions loaded on SDS-PAGE are shown in figure 15a–b. The homologues 33SSP, 40FJO, 55PSY and 69MCA

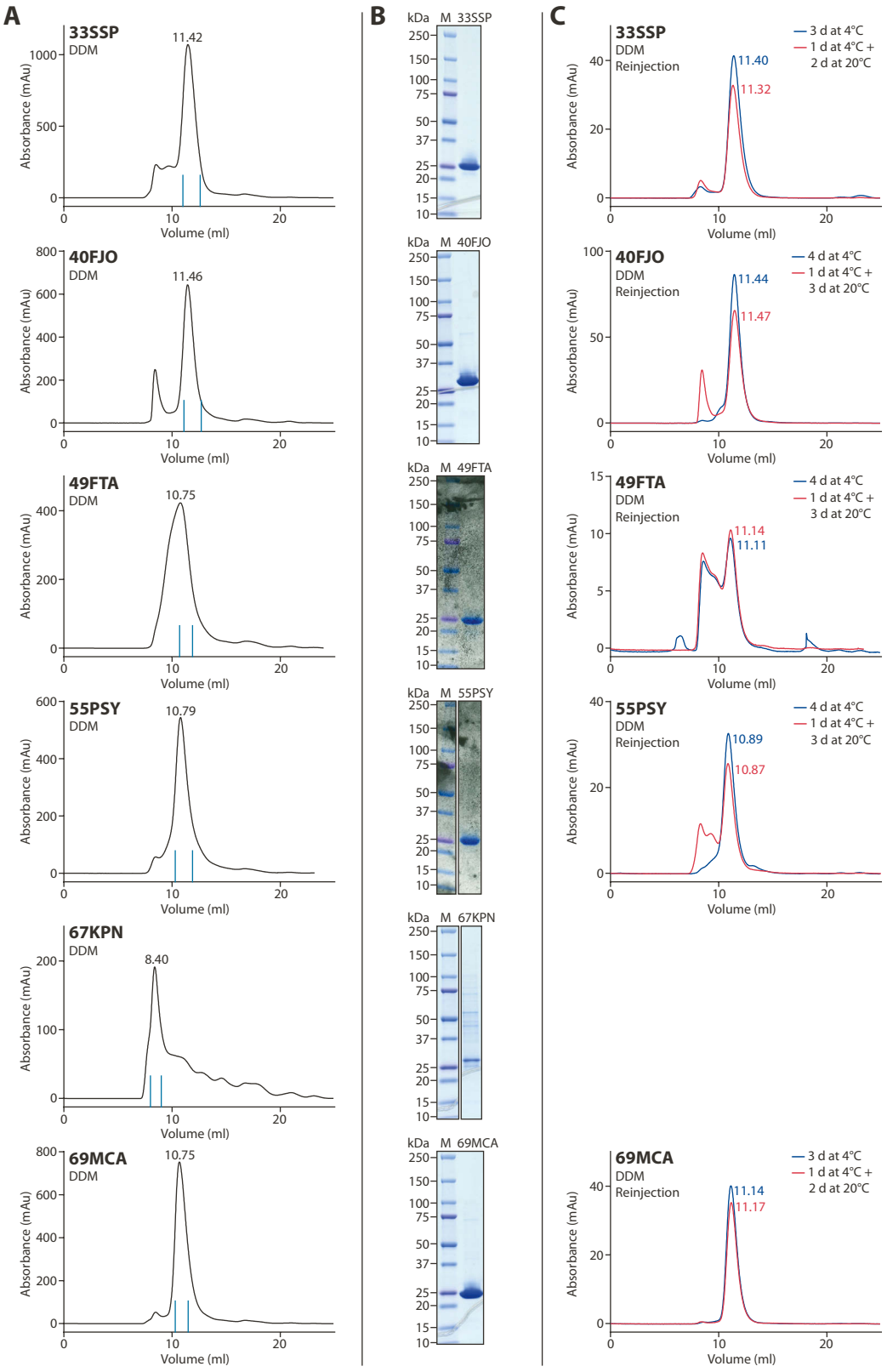


Figure 15 – Purification of homologues from the second round in DDM and test of stability. **(A)** SEC results. Blue lines on the bases of the chromatograms indicate the peak fractions. **(B)** SDS-PAGE analysis of the peak fractions. The gel containing 49FTA and 55PSY was contaminated with activated charcoal particles. **(C)** Parts of the peak fractions were reanalysed by SEC after being stored at the indicated temperatures for the indicated number of days at a concentration of 5 mg/ml. Traces are coloured blue for samples stored at 4°C and red for samples stored at 20°C.

all eluted as single monodisperse peaks. All four homologues eluted at volumes consistent with those observed for 3GAU and 12CPI. In contrast 69MCA and 55PSY eluted almost 0.7 ml earlier than 33SSP and 40FJO, similar as observed for the elution behaviour of 3GAU and 12CPI, where 3GAU eluted earlier than 12CPI. The difference could potentially be explained by the observation that 3GAU, 55PSY and 69MCA all had longer C-termini (Figure 11 in an earlier section), which were predicted to be disordered (data not shown).

The homologue 49FTA did not elute as a monodisperse peak and showed a large shoulder on the leading edge. The homologue 67KPN (from *Klebsiella pneumoniae*; identical to kpBEST, whose structure was later published by another group) started precipitating during proteolytic cleavage with HRV 3C protease. The supernatant injected on a size-exclusion column eluted in the void volume (Figure 15a).

For each homologue, a simple stability test was performed to test the long-term stability of the proteins (Figure 15c). 33SSP and 69MCA appeared to be stable for several days, both in the cold and at room temperature. For 40FJO and 55PSY a more significant amount of aggregation was observed when samples were stored at room temperature. However, since samples of 40FJO and 55PSY were stored longer than samples of 33SSP and 69MCA, direct comparisons should be taken with caution. Since similar to 67KPN also 49FTA appeared predominantly aggregated, both proteins were not investigated further, while 33SSP, 40FJO, 55PSY and 69MCA were all used for crystallisation studies.

33SSP, 40FJO and 69MCA all crystallised in various conditions. 55PSY, in contrast, neither crystallised in initial experiments nor in broader screening efforts carried out by Nathalie Cornillie, a master student working with me on the project. Focus was therefore shifted to the three homologues 33SSP, 40FJO and 69MCA for which crystals have been obtained in initial screening experiments.

2.2.4. Crystallisation of 33SSP

For 33SSP, besides the crystallisation in DDM, crystallisation was also attempted in shorter chain detergents n-Undecyl-β-D-maltoside (UDM) and DM. The latter was chosen since 33SSP was stable in DM, even when stored for two days at room temperature (Figure 16). However, with UDM and DM, only a few needle shaped crystals were observed, which were too small for diffraction analysis and thus not further investigated.

In DDM initial crystals of 33SSP were observed in several different conditions. Most crystals grew at low concentrations of PEG400 and only a few in drops with low concentrations of PEG4000. Crystals appeared over a range of pH values ranging from 5.5 to 9.4 and in different salts. One crystal diffracted better than all other crystals. As shown in figure 17, spots could be detected up to a resolution of approximately 4.5 Å, based on a visual inspection. A complete

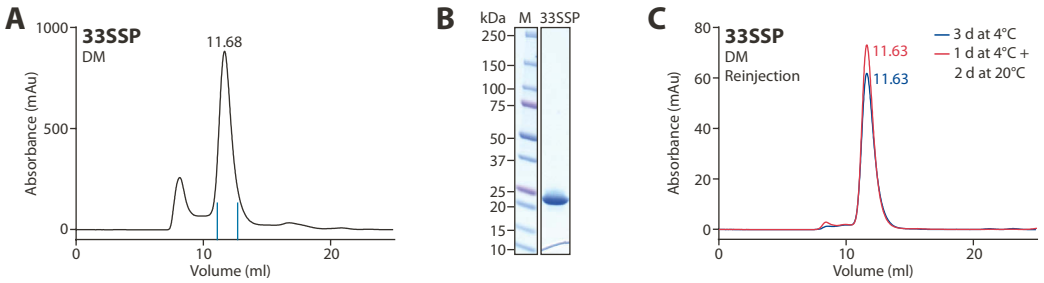


Figure 16 – Purification of 33SSP in DM and test of stability. **(A)** SEC result. Blue lines on the base of the chromatogram indicate the peak fraction. **(B)** SDS-PAGE analysis of the peak fraction. **(C)** Parts of the peak fraction were reanalysed by SEC after being stored at the indicated temperatures for the indicated number of days at a concentration of 5 mg/ml. Traces are coloured blue for samples stored at 4°C and red for samples stored at 20°C.

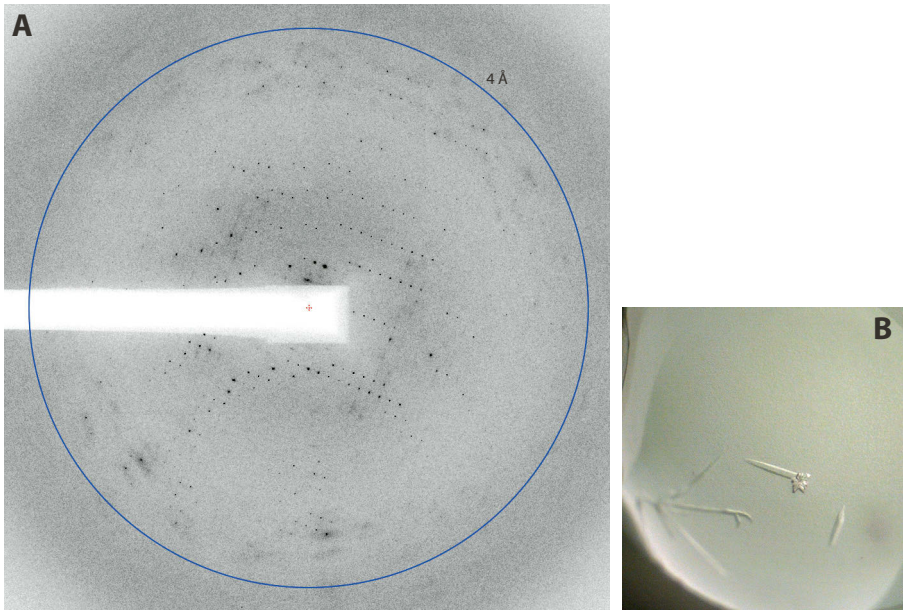


Figure 17 – Best diffracting crystal of 33SSP without any additives. **(A)** Observed diffraction pattern. Blue ring indicates the 4 Å resolution ring. **(B)** Picture of the drop in the crystallisation plate.

dataset was not collected, as the observed diffraction quickly deteriorated, possibly due to radiation damage. Therefore, the data was only processed for the first 25° of rotation. The data collection statistics are shown in table 5. The data extended to 4.8 Å, but was not complete due to the limited number of collected images. Interestingly, the crystallisation condition (500 mM KCl, pH 7.4, 15% PEG400) was the same as for the best diffracting crystal of 3GAU-C9. Attempts were made to improve the diffraction by variation of the pH, salt and PEG400 concentration and by increasing the volume of the setups. Even though the crystals did get bigger, their diffraction was not improved. Calcium added to the

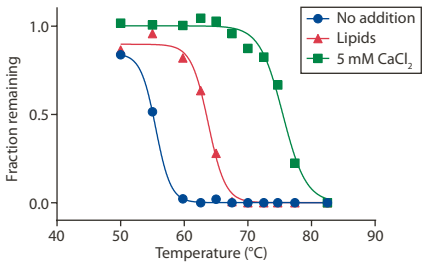


Figure 18 – FSEC based thermostability assay of 33SSP in UDM. The amount of non-denatured 33SSP was measured from the FSEC peak heights. Lipids were *E. coli* polar lipids added at a concentration of 0.5 mg/ml. Fraction remaining is the peak height normalised against the peak height of a control sample stored at 4°C.

drop did also not improve the diffraction. It was, however, observed that by slightly varying the PEG concentration and the pH value, crystals could be grown in UDM with similar diffraction properties. Using a simple thermostability assay, where small protein samples were incubated at different temperatures before being analysed by FSEC [80,81], it was observed that the addition of *E. coli* polar lipids increased the thermostability of 33SSP in UDM (Figure 18). An even larger effect was observed by the addition of calcium, even though calcium so far did not seem to make any difference in crystallisation.

Crystal	
Protein:	33SSP in DDM
Additive:	none
Condition:	500 mM KCl 50 mM HEPES pH 7.4 15% PEG400
Data collection	
Beamline:	X06SA, SLS
Wavelength:	0.920 Å
Space group:	P 1 2_1 1
Cell dimensions:	
a, b, c (Å)	113.94, 101.37, 114.68
α, β, γ (°)	90.00, 100.24, 90.00
Resolution (Å):	43.88 – 4.80 (4.88 – 4.80)
R_{merge} (%):	6.6 (30.7)
R_{meas} (%):	8.3 (39.4)
$CC_{1/2}$:	0.999 (0.874)
$\|I\|/\sigma(I)$:	6.7 (2.4)
Completeness (%):	24.31 (27.51)
Redundancy:	2.1

Table 5 – Details of crystallisation and data collection statistics for the best diffracting crystal of 33SSP without any additives. Numbers in parentheses indicate the highest resolution shell and its statistics. Data was only processed for a rotation of 25°, since the diffraction quality quickly deteriorated, possibly due to radiation damage.

New crystallisation screens were performed, using 33SSP purified in UDM with addition of lipids or lipids and calcium. Unlike in previous trials with 33SSP not containing additives, crystals were now observed in several drops containing lipids with several crystals growing in similar conditions as observed for 33SSP in DDM. In addition, a new condition was identified containing 200 mM $CaCl_2$, which grew in a pH range between 5.5 and 9.4. Like previously, crystals appeared in low concentrations of PEG400. The best diffracting crystal was found at a

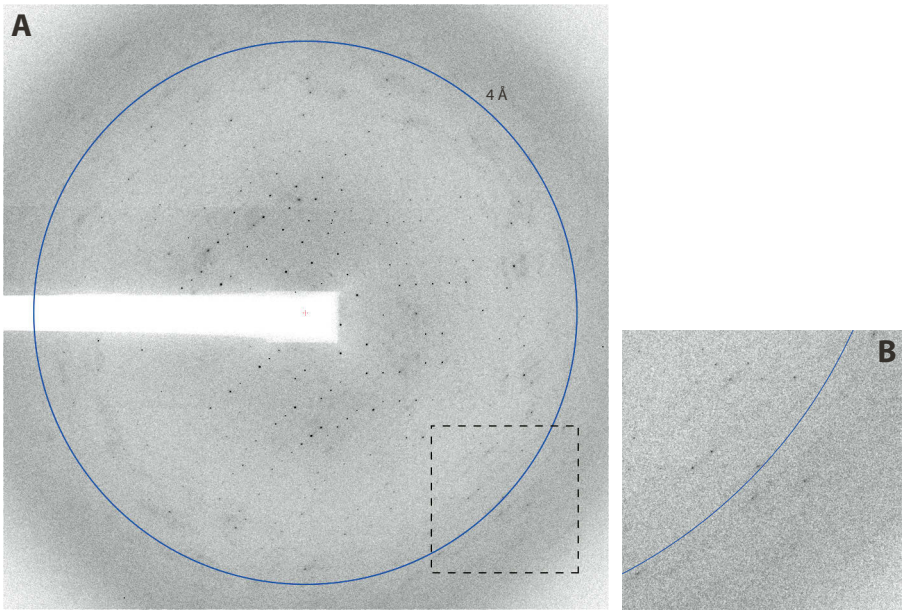


Figure 19 – Diffraction of 33SSP in UDM with the addition of lipids. 0.5 mg/ml *E. coli* polar lipids were added. **(A)** Observed diffraction pattern. Blue ring indicates the 4 Å resolution ring. The dashed line indicates the region which is enlarged in following panel. **(B)** Enlargement of the highlighted part of the diffraction image.

pH value of 7.4 and diffraction could be detected up to a resolution of approximately 4.0 Å (Figure 19). This appeared as a significant improvement over the previously described crystal without lipids, where spots could only be observed up to a resolution of 4.5 Å. The statistics from a partial dataset of this crystal are shown in table 6. The data extends to 4.7 Å and is thus only marginally better than the previously described dataset from the crystal grown without lipids.

Crystal	
Protein:	33SSP in UDM
Additive:	0.5 mg/ml *E. coli* polar lipids
Condition:	200 mM $CaCl_2$
	50 mM HEPES pH 7.4
	15% PEG400
Data collection	
Beamline:	X06SA, SLS
Wavelength:	1.000 Å
Space group:	P 1 2_1 1
Cell dimensions:	
a, b, c (Å)	106.51, 103.04, 114.80
α, β, γ (°)	90.00, 98.57, 90.00
Resolution (Å):	46.92 – 4.70 (4.78 – 4.70)
R_{merge} (%):	6.9 (39.1)
R_{meas} (%):	9.1 (51.8)
$CC_{1/2}$:	0.999 (0.789)
I/σ(I):	6.9 (2.5)
Completeness (%):	32.41 (34.41)
Redundancy:	1.9

Table 6 – Details of crystallisation and data collection statistics for the best diffracting crystal of 33SSP in UDM with lipids added. Numbers in parentheses indicate the highest resolution shell and its statistics. Data was only processed for a rotation of 30°, since the diffraction quality quickly deteriorated, possibly due to radiation damage.

Unfortunately, it proved to be difficult to reproduce crystals that diffracted equally well or better. Eventually, after optimising the pH, type of PEG, salt concentration and lipid type, a better diffracting crystal was identified. An example of the diffraction pattern is shown in figure 20. According to the data collection statistics shown in table 7, the data was complete to a resolution of 3.9 Å. Space group and unit cell dimensions were similar to previous datasets. Interestingly this crystal had been grown in DDM. Thus, despite being stable in UDM and DM, 33SSP did not only crystallise better in DDM, but the crystals grown in DDM also diffracted to higher resolution.

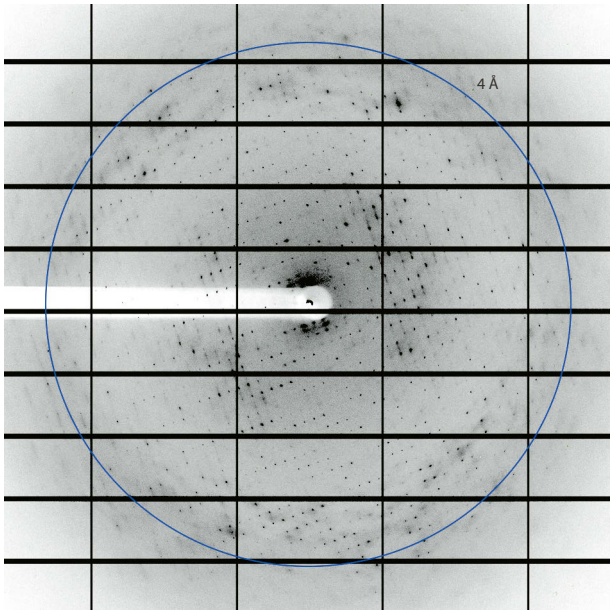


Figure 20 – Observed diffraction of 33SSP in DDM with high concentration of lipids and detergent. 1 mg/ml DOPC and 0.1% DDM were added. The crystallisation solution consisted of 200 mM $CaCl_2$, 50 mM HEPES pH 6.9 and 14% PEG300. Blue ring indicates the 4 Å resolution ring.

Even despite extensive trials it was not possible to obtain crystals of 33SSP diffracting higher than 3.9 Å, whereas most crystals grown in the same condition only diffracted to much lower resolution. As screening for additives did not bring any improvement, the method used for adding lipids to the purified protein was modified. Instead of adding a liposome solution directly to the purified protein, lipid mixtures in chloroform were added to small glass tubes. After evaporating the chloroform, concentrated protein and additional detergent was added directly to the glass tube along with a small stirrer bar. Typically, different amounts of lipid and detergent were tested. After stirring overnight, the non-solubilised lipids were removed by ultracentrifugation. This method, also known as HiLiDe [82], improved the reproducibility of the crystals, but still did not result in better diffracting crystals.

Crystal	
Protein:	33SSP in DDM
Additive:	1 mg/ml DOPC
Condition:	200 mM $CaCl_2$ 50 mM HEPES pH 6.9 14% PEG300
Data collection	
Beamline:	X06SA, SLS
Wavelength:	1.000 Å
Space group:	P 1 2_1 1
Cell dimensions:	
a, b, c (Å)	113.10, 101.04, 114.91
α, β, γ (°)	90.00, 99.96, 90.00
Resolution (Å):	49.37 – 3.90 (3.97 – 3.90)
R_{merge} (%):	7.7 (136.1)
R_{meas} (%):	8.3 (145.6)
$CC_{1/2}$:	0.999 (0.898)
$\|I\|/\sigma(I)$:	12.1 (2.2)
Completeness (%):	99.73 (99.40)
Redundancy:	7.7

Table 7 – Details of crystallisation and data collection statistics for the best diffracting crystal of 33SSP in DDM with lipids added. Numbers in parentheses indicate the highest resolution shell and its statistics.

2.2.5. Crystallisation of 69MCA

As previously mentioned, the homologue 69MCA was cloned from a thermotolerant organism. It therefore seemed likely that this protein would be more stable than other investigated homologues in different detergents. 69MCA was therefore from the beginning purified in both DDM and DM. Figure 21 shows the result of a purification in DM and a simple stability test performed with the purified protein. As shown in the figure, unlike some other investigated proteins, 69MCA is stable in DM, even at 20°C.

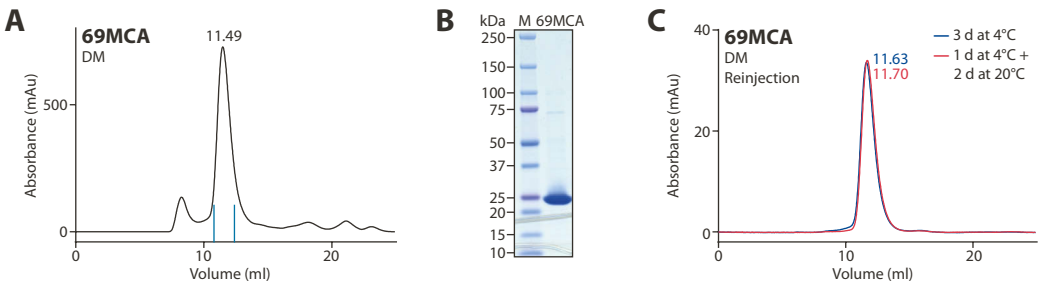


Figure 21 – Purification of 69MCA in DM and test of stability. **(A)** SEC result. Blue lines on the base of the chromatogram indicate the peak fraction. **(B)** SDS-PAGE analysis of the peak fraction. **(C)** Parts of the peak fraction were reanalysed by SEC after being stored at the indicated temperatures for the indicated number of days at a concentration of 5 mg/ml. Traces are coloured blue for samples stored at 4°C and red for samples stored at 20°C.

After extensive screening, crystals of 69MCA appeared in both detergents. In DDM the protein crystallised in different salts and at different pH values, in low concentrations of PEG400 or PEG4000. With DDM crystallisation was only attempted at 4°C. Unfortunately, none of the tested crystals diffracted to more than 8 Å. When purified in DM, no crystals were obtained at 4°C and crystals grown at 20°C had poor diffraction properties.

After discovering that lipids had a positive effect on the crystallisation of 33SSP, I attempted to crystallise 69MCA in the presence of lipids. For this purpose, 69MCA was purified in UDM and liposomes were added to the purified protein after concentration. Crystals appeared at both 4°C and 20°C, but the crystals were generally bigger when grown at 20°C and thus only those were tested. Unfortunately, none of these crystals showed promising diffraction behaviour.

Since 69MCA has a longer C-terminus than other homologues (Figure 11 in an earlier section), different C-terminal truncations were constructed and tested by small scale expression and FSEC (data not shown). Truncation of up to 7 residues, was tolerated and resulted in expression levels similar to the full-length protein. Crystallisation was attempted in DM, at both 4°C and 20°C. Although small crystals were observed, these were not large enough for diffraction studies.

2.2.6. Crystallisation of 40FJO

The initial crystallisation screens performed with 40FJO showed crystals in many different conditions. Crystals grew in a broad range from pH 5.5 to 9.4, but they seemed to grow best in the range from 6.5 to 8.4. Although most crystals were observed at low PEG400 concentrations, some also grew at low concentrations of PEG4000. Work with this homologue, described below, was carried out by master student Nathalie Cornillie, who performed the work as her master thesis project.

Based on the experience of crystallising 33SSP, lipids were included in the crystallisation trials. In this case lipids were added as liposomes after the protein had been concentrated to the final concentration used for crystallisation. As an additive, primarily DOPC was used, but *E. coli* polar lipids were also tested. As in the case of 33SSP, calcium was found to improve the thermostability of 40FJO (data not shown). 10 mM $CaCl_2$ was therefore included during screening. Crystals were observed in several different conditions. Crystals were frozen from several different conditions. Most diffracted to less than 8 Å, but two crystals showed better diffraction properties than the rest. The best diffracting crystal, grown in 1M NaCl at pH 8.4, resulted in visible diffraction spots up to a resolution better than 5 Å (Figure 22). Unfortunately, the diffraction turned out to be anisotropic and the data was only complete to a resolution of 6.2 Å (Table 8). The crystal system was found to be orthorhombic (space group P 2_1 2_1 2_1).

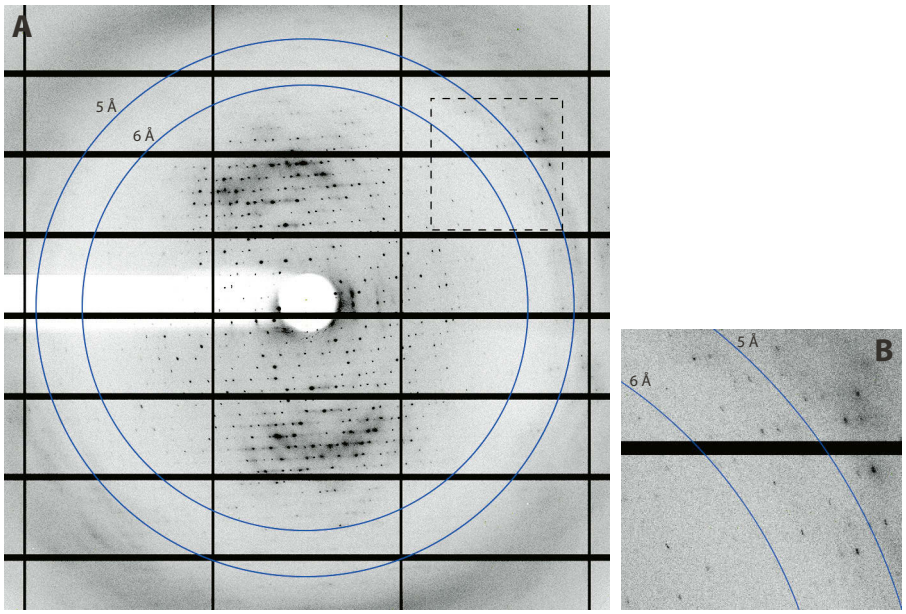


Figure 22 – Diffraction of 40FJO in DDM with lipids and calcium. 1 mg/ml DOPC and 10 mM $CaCl_2$ were added. **(A)** Observed diffraction pattern. Blue rings indicate the 5 and 6 Å resolution rings. The dashed line indicates the region which is enlarged in following panel. **(B)** Enlargement and contrast enhancement of the highlighted part of the diffraction image.

For comparison, the best diffracting 33SSP crystals were all monoclinic (space group P 1 2_1 1).

Another crystal, grown in 1M ammonium formate at pH 7.4, also showed diffraction spots extending to around 5 Å. The diffraction from this crystal was also anisotropic and the data was only complete to a resolution of 6.6 Å. The space group was found to be the same as for the first crystal (P 2_1 2_1 2_1), and unit cell dimensions were virtually identical (data not shown).

Several attempts were made to improve the quality of the crystals by refining the conditions. That did, however, not improve the diffraction. Other lipids were also tried without seeing any improvements.

Crystal	
Protein:	40FJO in DDM
Additive:	1 mg/ml DOPC, 10 mM $CaCl_2$
Condition:	1M NaCl
	50 mM Tris-HCl pH 8.4
	15% PEG400
Data collection	
Beamline:	X06SA, SLS
Wavelength:	1.000 Å
Space group:	P 2_1 2_1 2_1
Cell dimensions:	
a, b, c (Å)	118.85, 125.36, 180.68
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å):	49.65 – 6.20 (6.31 – 6.20)
R_{merge} (%):	6.7 (74.3)
R_{meas} (%):	7.4 (79.7)
$CC_{1/2}$:	0.995 (0.911)
$\|I\|/\sigma(I)$:	13.2 (2.6)
Completeness (%):	99.74 (100.00)
Redundancy:	7.0

Table 8 – Details of crystallisation and data collection statistics for the best diffracting crystal of 40FJO in DDM. Numbers in parentheses indicate the highest resolution shell and its statistics.

As with 33SSP, purification and crystallisation was also attempted using the detergent UDM. Crystals appeared in UDM and when PEG200 or PEG300 was used as precipitant, the crystals got significantly bigger than the crystals grown in PEG400 and DDM. Although the crystals

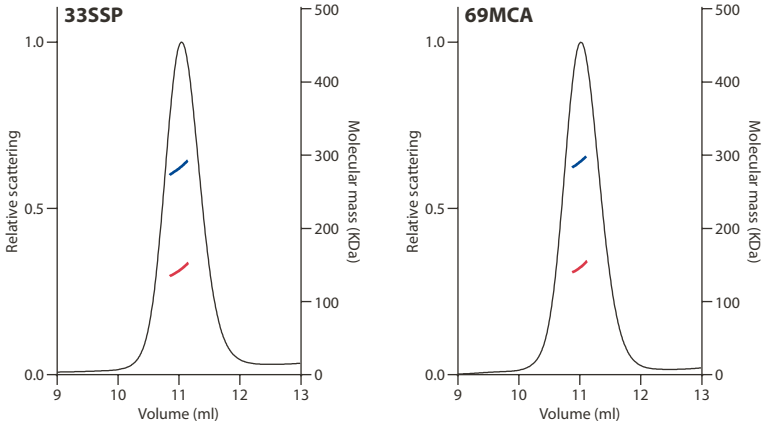
got bigger, the diffraction did not improve. Similarly, various additives were also tested, without any improvement.

In this case crystallisation was also attempted in the shorter detergent DM. Unlike 33SSP, 40FJO did crystallise in DM, but diffraction was not improved further.

2.2.7. Estimating the size of 33SSP and 69MCA

All purified bacterial bestrophins eluted from the size-exclusion columns at volumes that would suggest big, multimeric assemblies. At the time of these studies, no conclusive experiments had been performed to determine the exact oligomeric state of bestrophins. One study concluded that porcine BEST1 most likely forms a homodimer [33], based on ultracentrifugation experiments performed with porcine BEST1 extracted in Triton X-100. Another study found that overexpressed hBEST1 most likely forms either tetramers or pentamers [22]. For this study cells were co-transfected with two different BEST1 constructs that contained different epitope tags, and the stoichiometry was estimated based on co-immunoprecipitation.

To get a better understanding of the oligomeric state of bacterial bestrophins, multi-angle light scattering (MALS) experiments were performed with 33SSP and 69MCA purified in DDM. Using this method, it is possible to determine the mass of protein excluding detergent, as well as the mass of the whole-protein detergent complex. The results of this analysis are shown in figure 23. The molecular weights of the protein complexes without detergent, were determined to be a slightly below 150 kDa. Given the molecular weights of 33SSP and 69MCA monomers (34.7 and 32.2 kDa, respectively), the number of monomers per protein complex could be calculated. The calculated values were between 4 and 5 and it was thus not possible



Protein	33SSP	69MCA
MW (kDa) protein–detergent	284.6	292.5
MW (kDa) protein alone	144.7	148.2
MW (kDa) (calculated)	34.7	32.2
Stoichiometry	4.17	4.60

Figure 23 – Multi angle light scattering analysis of 33SSP and 69MCA. Chromatograms show the normalised scattering signals. Molecular weights of the protein–detergent complexes are shown in blue and the molecular weights of the proteins alone are shown in red. The table shows the measured, average molecular weights of the protein–detergent complexes as well as the molecular weights of the proteins complexes alone (without the detergent micelle). For comparison, the molecular weights of the 33SSP and 69MCA monomers (calculated from the amino acid sequences) are shown. The calculated stoichiometry is the ratio between the measured molecular weight of the protein complex and the calculated molecular weight of a monomer.

to say if the protein complexes were tetramers or pentamers. Some inaccuracy might be due to assumptions that were made for the protein–detergent complexes, regarding the refractive index increments of protein and detergent. It was assumed that the micelle consisted purely of protein and detergent. Other compounds, such as lipids, were not taken into account. The accuracy of the calculations might therefore have been affected by the presence of lipids.

2.2.8. Initial purification of homologues from the third round

Homologues identified in the third round of screening that performed well in the small-scale experiments, were all expressed in larger scale cultures and purified in DDM. Two of the homologues, 95FIN and 113FBA, performed well in this analysis. These homologues each resulted in a single, monodisperse peak eluting from the size-exclusion column (Figure 24a).

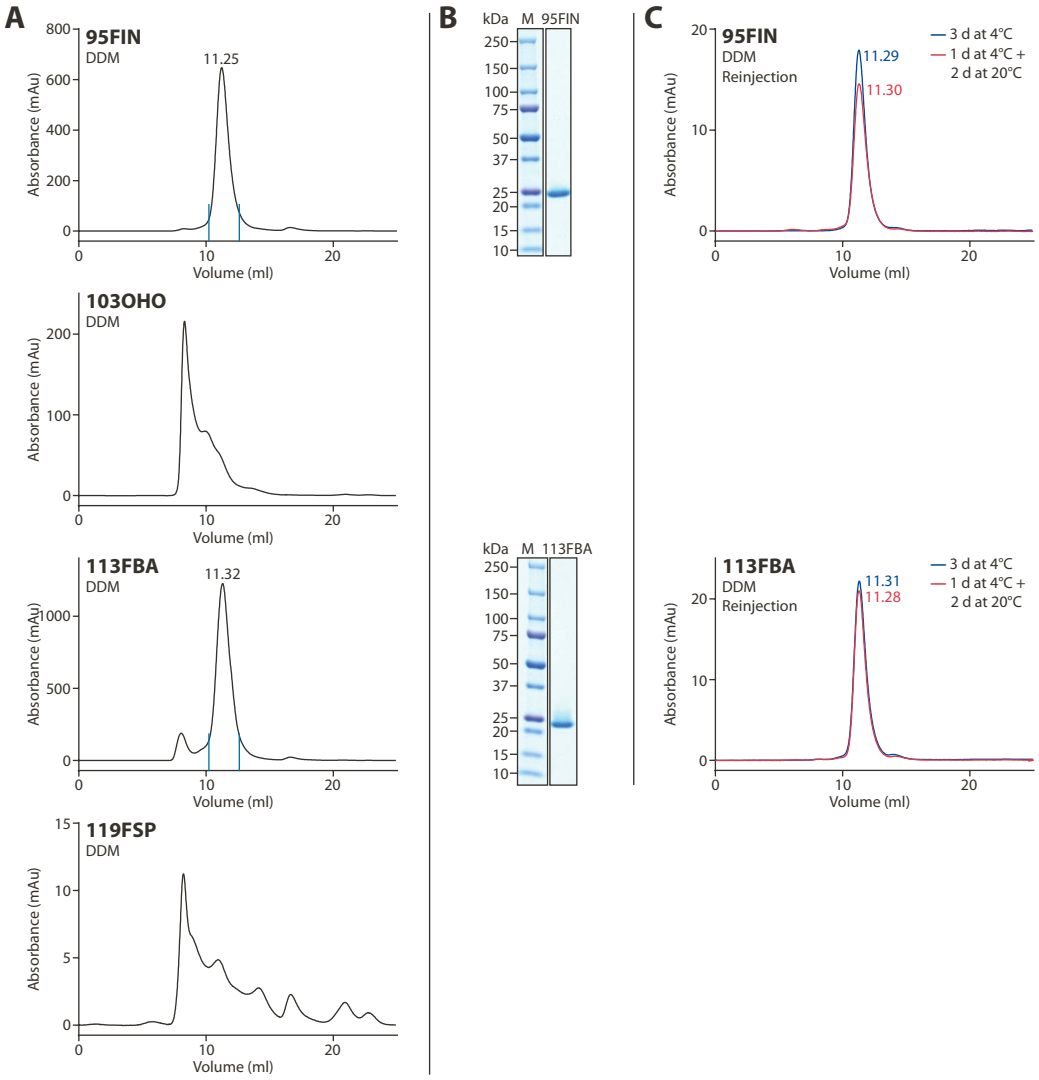


Figure 24 – Purification of homologues from the third round in DDM and test of stability. **(A)** SEC results. Blue lines on the bases of the chromatograms indicate the peak fractions. 103OHO and 119FSP were not analysed further. **(B)** SDS-PAGE analysis of the peak fractions. **(C)** Parts of the peak fractions were reanalysed by SEC after being stored at the indicated temperatures for the indicated number of days at a concentration of 5 mg/ml. Traces are coloured blue for samples stored at 4°C and red for samples stored at 20°C.

The protein appeared pure, as judged by SDS-PAGE (Figure 24b), and stable for several days at both 4°C and 20°C (Figure 24c).

Two other investigated homologues did not perform well, even though they each gave a single, monodisperse peak in the small-scale purification. The homologue 103OHO appeared to be aggregated, judging from the chromatogram (Figure 24a). The fractions containing the void volume were visibly green (data not shown), which indicated that the GFP-tag was not efficiently removed by HRV 3C protease, possibly because the protein was already aggregated at that point. This homologue was thus not investigated further.

The homologue 119FSP seemed to express at similar levels as other homologues, with a similar amount of protein eluting from the IMAC column (data not shown). However, after incubation with HRV 3C protease almost all protein was captured on the IMAC column, indicating that the protein was not cleaved. The small amount that did pass through the column was nonetheless analysed by SEC, but there was no single, monodisperse peak visible (Figure 24a). Therefore, this homologue was also not investigated further.

2.2.9. Crystallisation of 95FIN

Since the addition of lipids was beneficial for the crystallisation of 33SSP, I decided to crystallise 95FIN with lipids. For the initial crystallisation screens, more than 15 mg protein was purified. Two samples were prepared without lipids, either with $CaCl_2$ or EGTA added. Four samples were prepared with $CaCl_2$ and different lipid-detergent ratios. Finally, three samples contained EGTA in addition to different lipid-detergent ratios. Lipids (DOPC) were added using the HiLiDe method, which had worked well for 33SSP. Plates were set up exclusively at 4°C.

Within a week, crystals appeared in a variety of different salts, over a pH range from 5.5 to 9.4. Only a few tiny crystals appeared in PEG4000; all other appeared in low concentrations of PEG400 (typically 15%). In plates containing EGTA, crystals were only observed in few conditions and they were all very small. Overall crystals grew much better in conditions containing calcium.

20 conditions with promising crystals were selected for fine screening in larger drops with a gradient of PEG400. For each of the 20 conditions four different samples were tested. The first sample contained lipids and calcium, the second contained lipids without addition of calcium, the third contained calcium without addition of lipids and to the fourth neither calcium nor lipid was added.

In those plates crystals appeared both in drops with and without addition of calcium. Crystals were frozen from 10 different conditions but none of the tested crystals diffracted to a higher resolution than 7–8 Å. Since the homologues 33SSP and 40FJO had shown more promising diffraction properties, no further attempts were made to improve the diffraction of 95FIN.

2.2.10. Crystallisation of 113FBA

The initial crystallisation screening for 113FBA was done similarly to 95FIN, building on the experience obtained from 33SSP. The protein was purified in DDM. Three samples contained $CaCl_2$ in addition to lipid (DOPC) and detergent, which were both added using the HiLiDe method. These samples were used to set up plates at 4°C. Three additional samples of the protein were prepared. One contained EGTA and no added lipids, another contained $CaCl_2$ and no added lipids and the last contained $CaCl_2$ and lipids added from a liposome solution. These samples were used to set up plates at both 4°C and 20°C.

Like for the other homologues, crystals appeared typically within a week in plates that had been incubated at 4°C. Crystals could be observed in several different salts over a pH range from 5.5 to 9.4. In contrast to most of the other homologues, 113FBA appeared to crystallise well in low concentrations of both PEG400 and PEG4000. Ca^{2+} and the addition of lipids were beneficial for crystallisation.

Crystals were frozen from 10 different conditions and their diffraction was investigated. Although most crystals showed poor diffraction properties, one crystal grown in ammonium sulphate at pH 6.5 showed a well-defined diffraction pattern with spots up to a resolution of 4 Å (Figure 25). Processing the data, showed that the data was complete up to a resolution of 4.2 Å (Table 9). Like 40FJO, 113FBA crystallised in the P 2_1 2_1 2_1 space group, with very similar unit cell dimensions. Given the molecular weight of a monomer (34.4 kDa) and assuming that the protein would be present as a pentamer, these unit cell dimensions would most likely correspond to one pentamer per asymmetric unit and a solvent content of 66%. The high

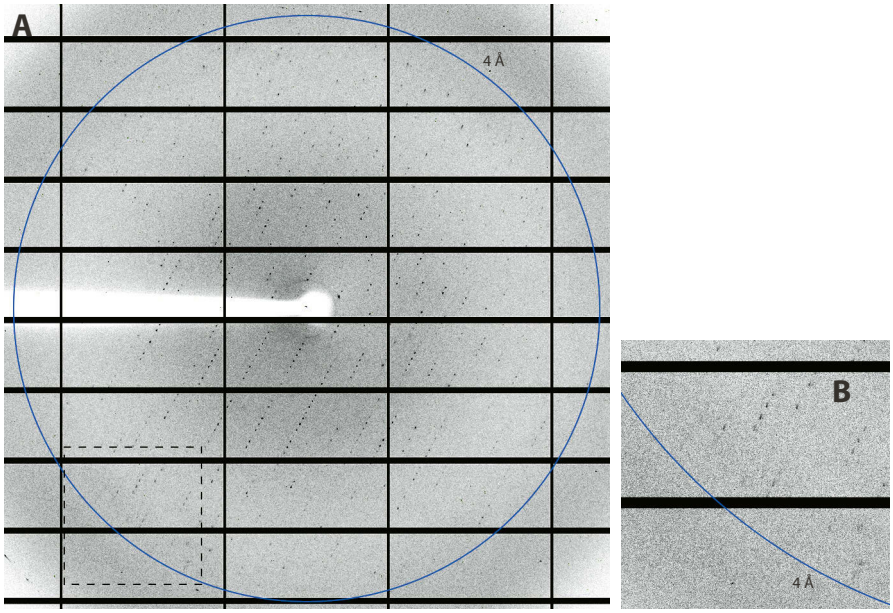


Figure 25 – Diffraction of 113FBA in DDM with the addition of high concentrations of lipid and detergent and 10 mM $CaCl_2$. **(A)** Observed diffraction pattern. The blue ring indicates the 4 Å resolution ring. The dashed line indicates the region which is enlarged in following panel. **(B)** Enlargement of the highlighted part of the diffraction image.

solvent content seemed reasonable, since membrane protein crystals include detergent micelles around the hydrophobic part of the protein.

Several attempts of improving this crystal form were made, but their diffraction was highly heterogeneous. Since after extensive screening all of the described attempts failed to yield crystals of a prokaryotic bestrophin homologue diffracting to higher resolution, all subsequent attempts were put on the improvement of crystals of 33SSP.

Crystal	
Protein:	113FBA in DDM
Additive:	0.15% DOPC, 0.75% DDM 10 mM $CaCl_2$
Condition:	200 mM $(NH_4)_2SO_4$ 50 mM ADA pH 6.5 15% PEG400
Data collection	
Beamline:	X06SA, SLS
Wavelength:	0.920 Å
Space group:	P 2_1 2_1 2_1
Cell dimensions:	
a, b, c (Å)	113.75, 124.66, 177.20
α, β, γ (°)	90.00, 90.00, 90.00
Resolution (Å):	49.67 – 4.20 (4.27 – 4.20)
R_{merge} (%):	10.0 (103.8)
R_{meas} (%):	10.4 (108.0)
$CC_{1/2}$:	0.999 (0.830)
$\|I\|/\sigma(I)$:	14.5 (2.4)
Completeness (%):	99.98 (100.00)
Redundancy:	13.2

Table 9 – Details of crystallisation and data collection statistics for the best diffracting crystal of 113FBA in DDM. Numbers in parentheses indicate the highest resolution shell and its statistics.

2.3. Targeting 33SSP with nanobodies

2.3.1. Nanobodies as crystallisation chaperones

The growth of well diffracting crystals of membrane proteins can be difficult. One cause for the poor quality of crystals is the poor stability of the target protein in the chosen detergent. This limitation can be addressed by using a different detergent, stabilising the protein with mutations or by co-crystallising the protein in complex with a stabilising ligand. In the case of 33SSP, different detergents and additives were tested. The introduction of stabilising mutations was, however, not attempted, since such mutations might result in a non-native conformation.

Another reason why membrane protein crystals can result in poorly diffracting crystals, can be that the detergent micelle limits the number of possible crystal contacts between proteins [78]. One solution frequently used to overcome this problem is the use of shorter chain detergents, which for 33SSP unfortunately did not improve the diffraction. Another option is to increase the accessible surface area by co-crystallising with a binding partner, a so-called crystallisation chaperone. Commonly used crystallisation chaperones for membrane proteins include antibody fragments, which have been used successfully for many years [83] and in many cases [84], or other protein scaffolds, such as monobodies [85] or designed ankyrin repeat proteins (DARPins) [86]. While monobodies and DARPins are always selected from synthetic libraries, antibody fragments used for crystallography are often selected from synthetic libraries or made from monoclonal antibodies. When selecting binders from a synthetic library, the screening can be fast and the proteins can subsequently be expressed recombinantly. In comparison, generation of monoclonal antibodies can be time consuming and costly, but does hold the advantage that the produced antibodies have been affinity matured by the immune system of the animal.

Antibody fragments can, however, also be generated by a combined approach, where a library is made from the blood cells of an animal that has been immunised with a target protein. Making such a library would be complicated for normal antibodies, since they consist of two separate chains, termed heavy and light chains. However, camelids and sharks also produce antibodies consisting only of heavy chains [87,88], in addition to the normal antibody types. By immunising a camelid, typically a llama or alpaca, with a target protein, a library can be made to select the variable part of such a heavy chain antibody [89]. This variable part known as VHH or a nanobody, is small (approximately 15 kDa) and can generally be recombinantly expressed in *E. coli*. Furthermore, it can often bind to epitopes that are not usually targeted by classical antibodies, such as clefts or active sites [90]. Nanobodies have by now been successfully used for crystallisation of many different proteins, including membrane proteins [91].

2.3.2. Selection of nanobodies

As nanobodies had previously proven to be critical to the crystallisation of several membrane proteins in our lab [76,77], I decided also to raise nanobodies against 33SSP, with the hope of producing better diffracting crystals. At the time, the nanobody technology had only recently been established at the University of Zurich, thanks to the work by Dr. Saša Štefanić (Vetsuisse Faculty, University of Zurich), Dr. Eric R. Geertsma (at the time a post-doc in our lab, currently at Institute of Biochemistry, Goethe University, Frankfurt) and others. For immunisation, an alpaca was injected four times with 33SSP, at two week intervals. 200 µg of 33SSP was used for each injection. After the last injection, a blood sample was collected and cDNA was generated from isolated lymphocytes. The genes encoding the variable fragments of the heavy chain antibodies were amplified from the cDNA and used to construct a phage display nanobody library. These steps were carried out by Yvonne Neldner in our group.

The library was based on a phagemid vector, i.e. a plasmid vector containing the f1 origin of replication from bacteriophage f1. The open reading frame consisted of the following elements: PelB-leader sequence, nanobody, 6×His-tag, EPEA-tag, HA-tag and finally the *gene III* from bacteriophage M13. An amber stop codon was placed between the EPEA-tag and the HA-tag, allowing an amber suppressor strain such as *E. coli* TG1 to occasionally fuse a nanobody to the phage coat protein pIII, encoded by *gene III*.

For the selection of nanobodies, 33SSP was immobilised on ELISA plates. Two different selections were performed. For one selection, 33SSP was coated directly on the plastic surface of the ELISA plate. For the other selection, 33SSP had been expressed with a site-specific biotinylation sequence [92] and biotinylated using the *E. coli* biotin ligase BirA, during the purification, prior to the size-exclusion chromatography step. This biotinylated 33SSP was then bound to an ELISA plate that had been coated with neutravidin. These two selection methods, which were termed «direct coating» and «neutravidin coating», were carried out in parallel, but kept strictly separate. For each selection method, a well was also prepared without 33SSP to be used as a negative control. Phages were added to all wells. After elution, the phages were used to infect *E. coli* TG1 cells. In order to produce new phage particles, the infected *E. coli* cells had to be superinfected with a helper phage, in this case M13KO7. Phage particles could then be recovered and used for a new round of selection.

Three rounds of selection were carried out. After each round, the number of phages was quantified in a dilution series of the eluted phages used to infect small amounts of TG1 cells. From the number of colonies and the dilution factor, the phage concentration could be calculated. The calculated concentrations are shown in table 10. The concentrations are not directly comparable between the different selection rounds and coating methods, since different amounts of input phages were used and since the amount of coated protein probably varied between the

two coating methods. However, by comparing the number of phages eluted from a well containing 33SSP with the number of phages eluted from the negative control, the degree of enrichment could be estimated. For the direct coating selection, a small amount of enrichment was observed already after the first round and after the second round this had increased to a 20-fold enrichment. A third round of selection did not lead to a higher enrichment. By comparison, the neutravidin coating selection reached a more modest enrichment of 5-fold after the second round, but was close to a 20-fold enrichment after the third round. Based on these results it was decided to analyse single clones from the second and third rounds of selection.

	Neutravidin coating		Direct coating	
	33SSP	**Negative control**	**33SSP**	**Negative control**
Round 1	5×10^3	5×10^3	2×10^4	5×10^3
Round 2	1×10^8	2×10^7	2×10^7	1×10^6
Round 3	7×10^7	4×10^6	9×10^6	5×10^5

Table 10 – Concentrations of phages eluted after the different rounds of phage display. The numbers, which show the number of phages per ml, were calculated from the number of colonies observed after infecting *E. coli* TG1 cells with a dilution series of the eluted phages. Columns labelled 33SSP show the output from wells coated with 33SSP. Columns labelled negative control, show the output from wells without 33SSP. Concentrations are shown for each of the two different selection strategies labelled "neutravidin coating" and "direct coating". In the first round, only one negative control was used, since this round used the same phage input for both selections.

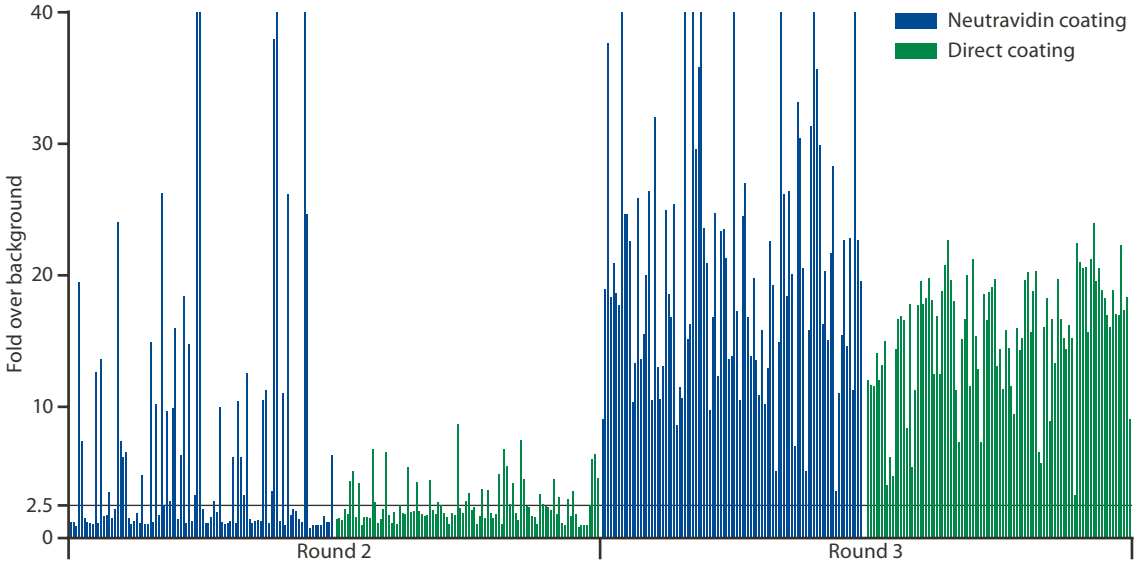


Figure 26 – Single-clone ELISA signals from the selection of nanobodies against 33SSP. Each coloured bar represents the signal from a single clone. Heights of the bars show how many times the signal was above the background signal, which was measured in wells without 33SSP but with nanobody. Some signals were more than forty times above the background, but have been reduced to the height of the plot in order to simplify the visualisation. Left and right halves show signals from ELISA performed after the second round and third round of phage display, respectively. Blue bars represent clones selected using the neutravidin coating procedure, whereas green bars represent those selected using the direct coating. A black line at 2.5× background represents the cut-off that was used to distinguish positive from negative signals. Heights are not directly comparable between different rounds and different coating techniques.

The single-clone analysis was performed using ELISA. From small-scale cultures, periplasmic extracts were prepared by subjecting the cells to an osmotic shock. The periplasmic extracts were then added to ELISA wells that had been coated with 33SSP, using the same coating methods as during the selection. For each clone, a negative control well without any 33SSP was prepared. This was used to exclude that the observed signal was due to unspecific binding. Figure 26 shows the result of the ELISA analysis. 96 clones were tested for each round and selection strategy. Each bar represents an ELISA signal divided by the background signal measured in the negative control well. The exact heights of the bars, cannot be compared between the results from round 2 and 3, since these were tested on different days. Similarly, the results from the two different selection strategies should not be directly compared, since the amount of coated protein would be expected to vary, given that the coating methods were different. It is, however, clear that far fewer clones from round 2 resulted in signals that were significantly above the background, when compared to round 3. In fact, all tested clones from round 3 gave positive signals, so it was decided not to perform more rounds of phage display.

Instead, plasmid was prepared from single clones and sequenced. Initially, only clones from round 3 were sequenced, with an equal number of clones from the two selections strategies. The clones from the direct coating strategy were found to almost exclusively contain the same nanobody sequence, whereas the clones from the neutravidin coating strategy showed significantly higher diversity. More clones were therefore sequenced from the neutravidin coating strategy. Additionally, clones were also sequenced from the second round of selection. In the end, it was counted how many times each nanobody sequence was observed. The result of this count is summarised in table 11. The nanobody sequence, NB33_1, that dominated the output from the 3rd round of the direct coating selection, was the only sequence observed

	Round 2		Round 3	
	Direct coating	Neutravidin coating	Direct coating	Neutravidin coating
NB33_1	16	11	45	32
NB33_2			1	5
NB33_3		1		4
NB33_4				3
NB33_5		5		6
NB33_6		1		3
NB33_7				1
NB33_8		7		8
NB33_9				1
NB33_10				1
NB33_11		3		9
NB33_12				1
NB33_13				5
NB33_14		1		
NB33_15			1	
NB33_16		1		
NB33_17				3
NB33_18		1		2
NB33_19				2
NB33_20		2		2
NB33_21				1
NB33_22				1
NB33_23				1
Total	16	33	47	91

Table 11 – Count of identical nanobody sequences. The table shows the number of times that each nanobody was observed during sequencing of single clones from round 2 and 3, using the two different selection strategies "direct coating" and "neutravidin coating". For the purpose of counting, nanobodies were considered identical, if they had identical complementarity-determining regions (CDRs) with the surrounding regions only showing minor differences.

in the output of the 2nd round of the same selection. Additionally, this nanobody accounted for approximately a third of the sequences from the neutravidin coating selection, from the 2nd as well as the 3rd round. No other nanobody was observed as frequently, although some were observed several times. 10 of the nanobodies were only observed once out of a total of 187 sequences. It seemed likely that more unique sequences could potentially be discovered if more clones were to be analysed, but it was decided to test the expression of the identified nanobodies, before analysing additional clones.

Of the 23 selected sequences, 22 were successfully cloned to an expression vector. One nanobody, NB33_21, could not be cloned despite several attempts and was therefore not used for further analysis. A multiple sequence alignment was made in order to characterise the sequence diversity of the selected nanobodies (Figure 27). As expected, the most sequence variation was observed in the three complementarity determining regions (CDRs), which were identified by comparing the sequences of the surrounding regions to those of llama nanobodies [93]. Also, as expected from llama nanobodies, the greatest length variation was seen

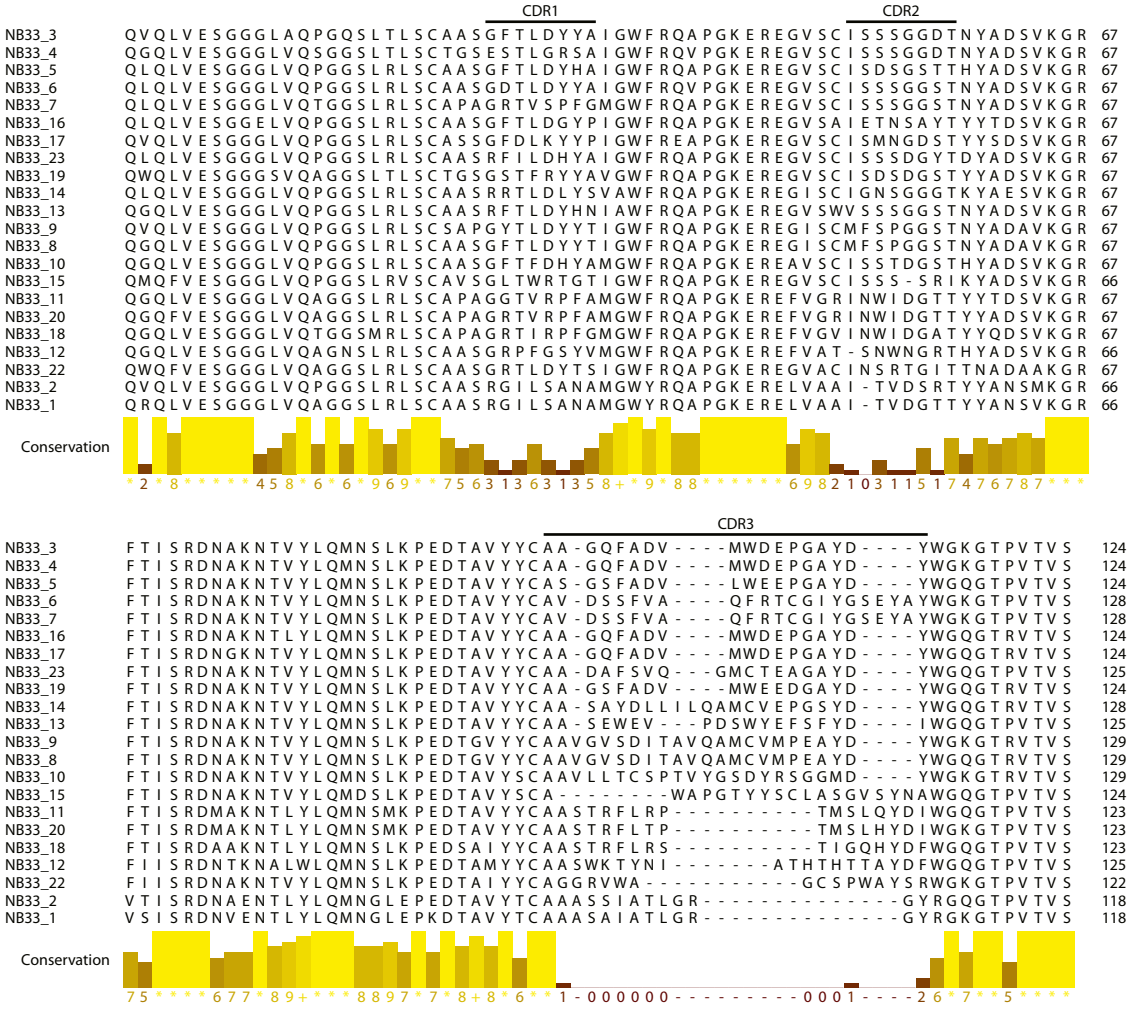


Figure 27 – Multiple sequence alignment of identified nanobodies. One nanobody (NB33_21) is not included, since it could not be successfully cloned. The three complementarity-determining regions (CDRs) are labelled as CDR1–3. Residues encoded by the plasmid vector are not included.

in CDR3, where the length varied between 13 and 23 amino acid residues. By comparison, CDR1 had a length of 8 amino acid residues in all identified nanobodies, whereas CDR2 varied between 7 and 8 amino acid residues. 8 nanobodies contained a cysteine residue in CDR3 in addition to a cysteine residue before CDR2, indicating that an extra disulphide bridge could be present, like it has been suggested for some llama nanobodies [94].

From the 22 selected nanobodies, few were eliminated from further investigation. Nanobody NB33_2 was not persued, since its sequence was very similar to that of NB33_1. The two nanobodies had the exact same sequence in CDR1, two amino acid residue substitutions in CDR2 and one in CDR3. Since NB33_1 was found in many more clones, it was chosen instead of NB33_2. The CDRs of NB33_9 were found to be virtually identical to those of NB33_8, with the only difference being a phenylalanine to tyrosine substitution in CDR1. Since NB33_8 was observed in more clones than NB33_9, NB33_8 was kept for further analysis. More nanobodies could have been excluded in this way, since for example NB33_20 and NB33_11 also showed very high similarity. However, they were both included since, unlike NB33_9, they had been identified in several clones during the sequencing. Other similar pairs of nanobodies, such as NB33_3 and NB33_4 or NB33_6 and NB33_7, were also included in the following studies.

2.3.3. Expression, purification and complex formation

All selected nanobodies were cloned into an *E. coli* expression vector resulting in the fusion of maltose binding protein (MBP) to the N-terminus of the protein, since this had previously been shown by others to result in consistently high expression levels [95]. The vector additionally included the PelB leader sequence for periplasmic targeting, a 10×His-tag to aid in purification and a HRV 3C cleavage site to remove the 10×His-MBP fusion by proteolytical cleavage. As for the expression of the bacterial bestrophins, the expression was under the control of the L-arabinose inducible P_{araBAD} promoter.

Each nanobody was initially expressed in 1.2 l of *E. coli* culture and purified using IMAC and SEC. The peak fractions and some additional fractions were further analysed by SDS-PAGE. The results are shown in figure 28. Most nanobodies resulted in a single monodisperse peak, with only smaller amounts of protein eluting at the void volume. An example of such a nanobody is NB33_1, which also resulted in a relatively high peak, indicating a good yield. However, when judging the yields based on UV-absorption, the extinction coefficients had to be considered, since they varied significantly between different nanobodies. For NB33_1 the calculated absorbance of a 1 mg/ml solution equalled 1.15, whereas for NB33_13 the value was 2.7. Some nanobodies, such as NB33_10 had a very high void peak compared to the main protein peak. As shown on the gels, these peaks generally contained very little protein migrating at the size expected for a nanobody (around 15 kDa). Most of the analysed void fractions

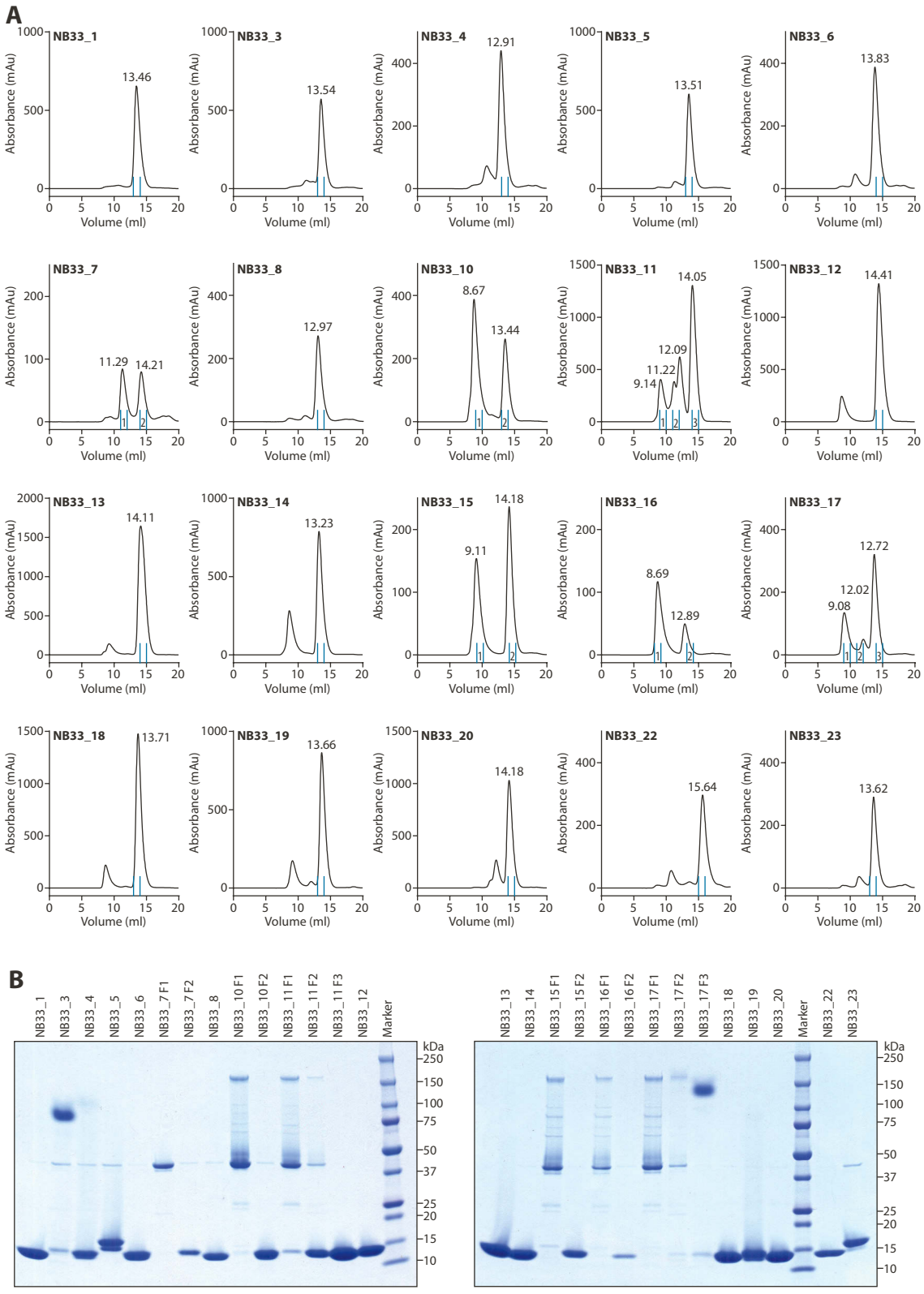


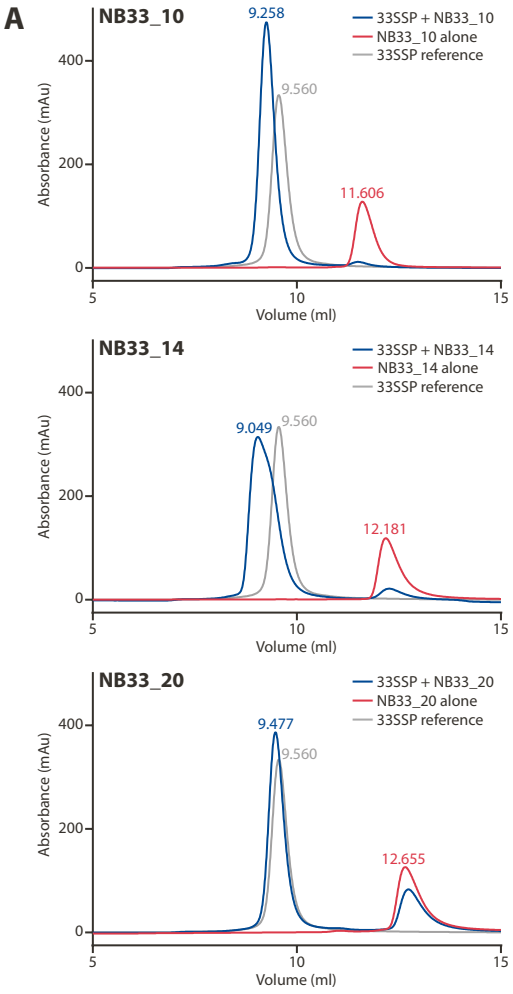
Figure 28 – Expression and purification test of selected nanobodies. 20 different nanobodies were purified from 1.2 l *E. coli* culture. **(A)** Size-exclusion chromatograms from the purification of different nanobodies on a Superdex 75 column. Blue lines on the bases of the chromatograms indicate the peak fractions, which were analysed by SDS-PAGE. Fractions are numbered for the cases where more than one fraction from a purification were analysed. Elution volumes for the peaks are shown as numbers next to the peaks. **(B)** SDS-PAGE analysis of peak fractions and a few other selected fractions. Samples were not boiled before loading on the gel.

instead contained a quite dominant band between 37 and 50 kDa, which would be consistent with the molecular weight of the cleaved off 10×His-MBP-tag. The presence of these void peaks was not problematic, since they were well separated from the main nanobody peaks.

Two nanobodies, NB33_3 and NB33_17, showed strong bands at high molecular weight on SDS-PAGE gels in addition to faint bands at the expected molecular weights. Before, the samples had not been boiled, to make it easier to compare these gel samples to complexes that also contain membrane protein, which generally cannot be boiled for SDS-PAGE. Since both nanobodies eluted from the size-exclusion column as well defined, monodisperse peaks, they were both included in the further analysis.

Four nanobodies, NB33_7, NB33_15, NB33_16, and NB33_22, were poorly expressed and thus not included in further analysis. For other nanobodies the yields varied between 0.5 mg per litre of culture for the worst expressing nanobody (NB33_17) to 2.8 mg per litre for the best expressing one (NB33_13). By comparison, NB33_1 which was the nanobody most frequently found in the selection had a yield of 1.9 mg per litre.

Next, it was tested if the 16 remaining nanobodies could form stable complexes with 33SSP. This would be expected if they were binding with high affinity. For this purpose, each nanobody was mixed with purified 33SSP in a 1:1 molar ratio. Each sample was then analysed by SEC. As controls, samples containing either nanobody or 33SSP alone were injected on the column. The control samples used the same amount of nanobody or 33SSP as the analysed samples. Figure 29a shows as an example the results obtained for the nanobodies NB33_10, NB33_14 and NB33_20. The complex between 33SSP and NB33_10 resulted in a peak that was visibly shifted to a smaller elution volume. The peak remained symmetrical and monodisperse, and almost all the added nanobody appeared to be bound to 33SSP as evidenced by a very small peak of free nanobody. The mixture of 33SSP and NB33_14 also resulted in a large peak shift, but in this case, the peak was broader and asymmetric, with a shoulder on the trailing edge. The peak shift for this complex was much bigger than that for the 33SSP-NB33_10 complex. The complex between 33SSP and NB33_20 gave only a very small peak shift and in this case a large fraction of the nanobody remained unbound. Figure 29b lists the elution volumes for all investigated complexes. All resulted in a peak shift, but the size of the shift varied from nanobody to nanobody. Based on a calibration curve for the size-exclusion column, the expected shifts in elution volumes were calculated for different binding stoichiometries (Figure 29c). All nanobodies, except NB33_1 and NB33_14, eluted within the expected range. When comparing the actual elution volumes with the expected values, it seemed clear that the binding stoichiometry varied between the different nanobodies. However, since the expected differences were relatively small, the exact binding stoichiometries could not be determined.



B

Injected sample	Elution volume (ml)
33SSP + NB33_1	9.191
33SSP + NB33_3	9.280
33SSP + NB33_4	9.344
33SSP + NB33_5	9.282
33SSP + NB33_6	9.325
33SSP + NB33_8	9.464
33SSP + NB33_10	9.258
33SSP + NB33_11	9.427
33SSP + NB33_12	9.473
33SSP + NB33_13	9.309
33SSP + NB33_14	9.049
33SSP + NB33_17	9.356
33SSP + NB33_18	9.429
33SSP + NB33_19	9.350
33SSP + NB33_20	9.477
33SSP + NB33_23	9.334

C

Nanobodies per pentamer	Expected elution volume (ml)
0	9.57 *
1	9.50 **
2	9.43 **
3	9.36 **
4	9.29 **
5	9.23 **

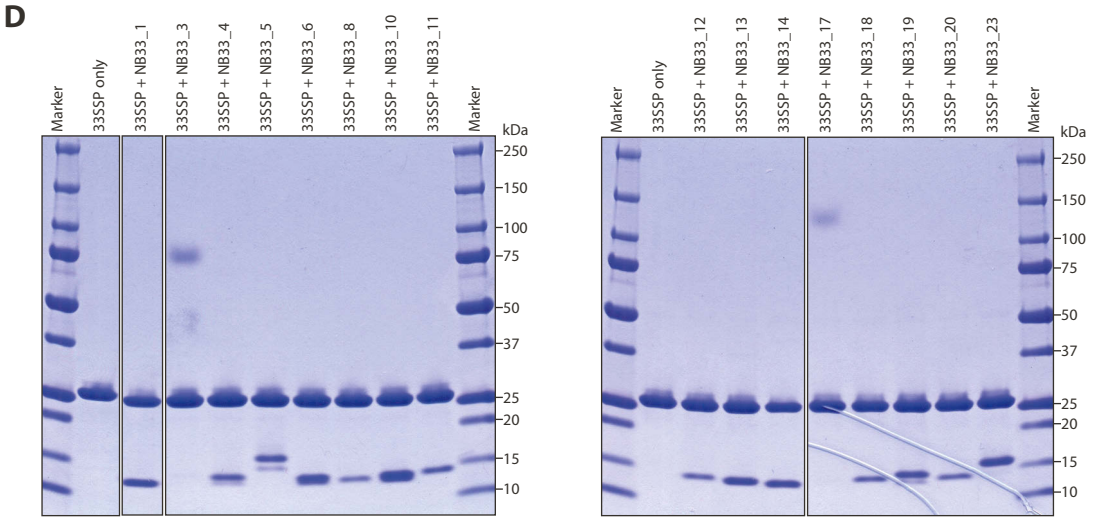


Figure 29 – Test of complex formation between purified nanobodies and 33SSP. 16 successfully purified nanobodies were mixed with purified 33SSP in a 1:1 molar ratio. **(A)** Examples of chromatograms obtained from SEC analysis of nanobody complexes. Blue lines show the chromatogram obtained when 33SSP and the respective nanobodies were injected together. Red lines show the same amount of nanobody injected alone. As a reference, the grey lines show 33SSP injected alone (the same chromatogram is shown in all three plots). **(B)** Elution volumes of all tested complexes. **(C)** Expected elution volumes. *Average of three SEC runs. **Calculated based on an average NB molecular weight of 13.8 kDa, using a calibration curve made for the size-exclusion column. **(D)** SDS-PAGE analysis of peak fractions. Samples were not boiled before loading on the gel.

For all 16 tested nanobodies, SDS-PAGE was used to confirm that the peak fractions contained both nanobody and 33SSP. As shown in figure 29d, all 16 nanobodies formed complexes with 33SSP. As observed previously (Figure 28b), two of the nanobodies (NB33_3 and NB33_17) migrated abnormally on the gel, despite behaving well in the SEC analyses. The intensities of the different nanobody bands varied markedly, but it was unclear if this was due to some nanobodies staining better or due to different binding stoichiometries. Interestingly, the nanobodies NB33_8, NB33_11, NB33_12, NB33_18 and NB33_20 resulted in relatively faint bands on the gels and only caused small shifts in the elution volumes during the SEC analyses (Figure 29b). Furthermore, during the SEC analyses a large amount of each of these nanobodies remained unbound (data only shown for NB33_20, figure 29a). It thus seemed likely that these nanobodies were binding to 33SSP at a low stoichiometric ratio, compared to the other tested nanobodies.

2.3.4. Crystallisation of 33SSP–nanobody complexes

For crystallisation screening, complexes of 33SSP and nanobodies were purified by SEC. Only the nanobodies NB33_1, NB33_5, NB33_6, NB33_8, NB33_11, NB33_12, NB33_13, NB33_14 and NB33_23 were tested, due to time restrictions. All complexes were purified using a Superdex 200 column. Most complexes eluted between 10.6 ml and 10.9 ml, except for complexes with NB33_1 or NB33_14, which eluted at approximately 9.6 ml (data not shown). Given such a large shift in elution volume, it seemed likely that these nanobodies were causing dimers of 33SSP-multimers to form. Interestingly, when NB33_1 or NB33_14 were purified without 33SSP, they eluted at similar volumes to the other nanobodies, suggesting that the nanobodies alone were not forming dimers (data not shown).

Crystallisation plates were set up both with and without the addition of lipids (using the previously described HiLiDe method). Since 33SSP alone had been found to crystallise predominantly in PEG400, the crystallisation screen, previously used for crystallising different bestrophin homologues, was modified slightly. Where the previous version of the screen contained both PEG400 and PEG4000, the new version of the screen contained only PEG400, in 8 different concentrations. The lowest concentration of PEG400 was 10% in the new version compared to 15% in the old versions, and the concentration was increased in 5% steps compared to the previous 10% steps.

For the 33SSP–NB33_6 complex, no crystals were observed. For 33SSP–NB33_8, 33SSP–NB33_12 and 33SSP–NB33_23 some crystals were observed, but their diffraction properties were not investigated further, because the crystals were small. For these complexes, crystals mainly appeared at PEG400 concentrations between 15% and 25%, both in the presence and absence of lipids.

For 33SSP–NB33_11 and 33SSP–NB33_13, the diffraction of crystals from several different conditions were tested, most of which were grown in the presence of lipids. Unfortunately, all tested crystals diffracted to very low resolution (typically less than 10 Å). 33SSP–NB33_5 crystals were only observed in the absence of lipid but none of them showed promising diffraction and in the best cases, diffraction spots were visible to a resolution of approximately 7 Å.

Unlike the previously tested complexes, 33SSP–NB33_14 almost exclusively crystallised in drops containing 10% PEG400. A few crystals did appear in some conditions with 15% PEG400, but the majority of the crystals would probably not have been observed in a screen containing 15% PEG400 as the lowest concentration. 33SSP–NB33_14 crystallised in a wide range of different conditions, both in the presence and absence of lipids. However, only crystals grown in the presence of lipids were tested, since they were generally bigger and had more regular shapes. Most of these crystals diffracted significantly better than what was observed for the previously tested complexes. The highest resolution where diffraction spots could be observed was generally between 7 Å and 4 Å, depending on the crystal. Further attempts to optimise the diffraction of this complex were not made, as the efforts were focused on another complex.

Complexes of 33SSP and NB33_1 also crystallised in many different conditions, both with and without lipids. Like for 33SSP–NB33_14, many crystals were found in drops containing 10% PEG400. Although crystals grown without lipids diffracted only to low resolutions, a big improvement was observed for crystals grown in the presence of lipids. For several crystals, diffraction spots could be observed up to a resolution of approximately 4 Å. A particularly promising condition consisted of 10% PEG400 at pH 6.5. The condition contained no salt, so the final drop only contained the salt that was included together with the protein in the SEC buffer. This crystal also showed diffraction spots up to a resolution of 4 Å, but the diffraction pattern looked less anisotropic than some of the other tested crystals and the diffraction pattern consisted of well-defined spots.

The crystallisation condition was optimised by varying the concentration of PEG and the pH value. Different salt concentrations were tested by reducing the salt concentration of the protein solution after the SEC. Additionally, some drops were set up with either $CaCl_2$ or EGTA. Crystals were also prepared using either PEG300 and PEG200 as precipitant and different cryoprotection strategies were tested. Eventually, one crystal grown at pH 6.5 in 13% PEG300 with the addition of 2 mM EGTA, was found to diffract far better than what had previously been observed. Figure 30 shows the observed diffraction pattern, extending to approximately 3 Å. Also shown in the figure is a picture of the crystal, which was grown in 1 µl sitting drops (final drop volume) and was estimated to measure approximately 150–200 µm in the longest direction.

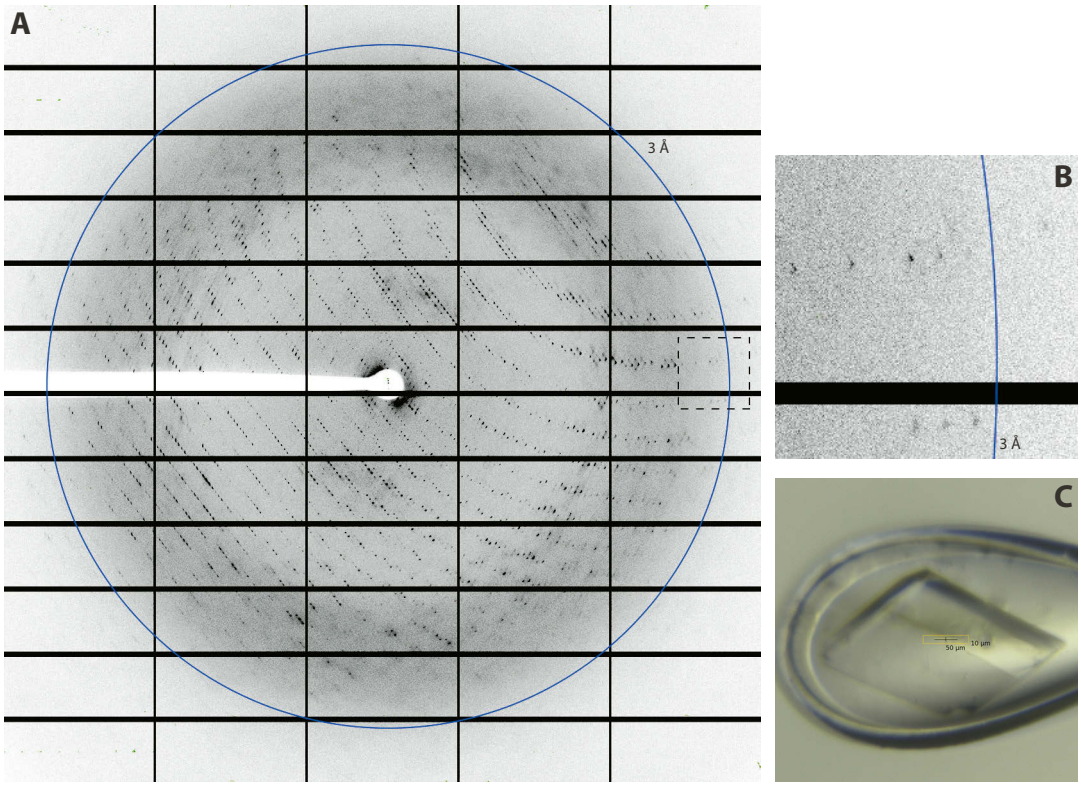


Figure 30 – Diffraction of 33SSP–NB33_1 in DDM with the addition of high concentrations of lipid and detergent and 2 mM EGTA. **(A)** Observed diffraction pattern. The blue ring indicates the 3 Å resolution ring. The dashed line indicates the region which is enlarged in following panel. **(B)** Enlargement of the highlighted part of the diffraction image. **(C)** Picture of the crystal in the loop, mounted at the beamline.

2.4. Structure determination of 33SSP in complex with a nanobody

2.4.1. Data collection and data processing

For the data collection of a 33SSP–nanobody complex, several parameters had to be taken into account in order to collect a complete dataset. Previous indexing of another crystal from the same condition had indicated that the crystal system could be triclinic. Since the only symmetry related reflections from a triclinic crystal are the Friedel pairs, a minimum of 180° rotation is required to get a complete dataset. However, since the detector used for the data collection (Pilatus 6M) contains gaps between the detector tiles, some reflections generally are not measured since they fall in a gap region. Using a larger rotation range can ensure that at least one half of the Friedel pair is measured in such cases. Also, a larger rotation range increases the redundancy, which can lead to a higher accuracy of the reduced structure factors [96]. On the other hand, a too large rotation range can lead to excessive radiation damage. Therefore, a full rotation of 360° was chosen.

With single-photon-counting detectors such as the Pilatus 6M, data quality can sometimes be improved when the rotation range captured for each image, is approximately half of the crystal mosaicity [97]. Based on previous indexing of another crystal from the same condition, the mosaicity had been estimated to be approximately 0.35°. A rotation range of 0.2° per image was chosen for the data collection. The detector was placed at a distance of 550 mm, since this distance would allow complete data to be collected up to a resolution extending slightly beyond 3 Å.

The collected data was indexed, scaled and merged with the XDS software package [98]. The data collection statistics are summarised in table 12. The space group was determined automatically using the program POINTLESS [99]. As expected from analysis of previous crystals, the space group was found to be P1. Manual inspection of the output files from XDS gave the same result. Data was included up to a resolution of 3.1 Å. This included relatively weak reflections, since doing so has been shown to generally improve the model quality [100]. The mosaicity reported by XDS was only 0.23° and

Crystal	
Protein:	33SSP–NB33_1 in DDM
Additive:	0.15% *E. coli* polar lipids, 0.75% DDM 2 mM EGTA
Condition:	*no additional salt* 50 mM ADA pH 6.5 13% PEG300
Data collection	
Beamline:	X10SA, SLS
Wavelength:	1.00002 Å
Space group:	P 1
Cell dimensions	
a, b, c (Å):	99.08, 183.97, 211.88
α, β, γ (°):	107.70, 92.18, 103.15
Resolution (Å):	48.86 – 3.10 (3.15 – 3.10)
R_{merge} (%):	9.6 (106.7)
R_{meas} (%):	11.3 (125.8)
$CC_{1/2}$:	0.997 (0.632)
$\|I\|/\sigma(I)$:	8.7 (1.2)
Completeness (%):	98.43 (98.33)
Redundancy:	3.5

Table 12 – Details of crystallisation and data collection statistics for the best diffracting crystal of 33SSP–NB33_1 in DDM. Numbers in parentheses indicate the highest resolution shell and its statistics.

thus lower than expected. An ideal data collection strategy would therefore have involved an even smaller oscillation range. However, the difference between an oscillation range of half the mosaicity and an one equalling the mosaicity were reported to be very small [97], so most likely a slightly smaller oscillation range would not have made much of a difference.

For the space group P1, one unit cell is known to contain one asymmetric unit. In order to estimate the number of biological units for each asymmetric unit, the Matthews coefficient (asymmetric unit volume per unit of protein molecular weight) was calculated for different numbers of biological units. For the calculation, it was assumed that 33SSP forms a pentamer and that five nanobodies would be bound to each pentamer, resulting in a molecular weight of 237 kDa for the 33SSP–NB33_1 complex. The Matthews coefficients can be compared to the values expected from the known protein structures present in the Protein Data Bank [101]. Since membrane protein crystals, as previously discussed, typically have a high solvent content because of the detergent micelles, the most likely numbers of biological units were considered to be 5, 4 or 3, corresponding to Matthews coefficients of 3.00 $Å^3$/Da, 3.75 $Å^3$/Da or 5 $Å^3$/Da, respectively. Based on a typical protein density of 1.35 g/cm^3 [102], these Matthews coefficients were expected to correspond to solvent contents of 59%, 67% or 75%, respectively. As previously discussed, the SEC elution profiles of 33SSP–NB33_1 suggested that the nanobody NB33_1 caused the formation of dimers of 33SSP oligomers. Although it was not known if such dimers would also be present in the crystal, it was still considered likely. Therefore, the number of biological units per asymmetric unit was assumed to most likely be 4, since this would correspond to two of these dimers.

2.4.2. Molecular replacement

In macromolecular crystallography, the next challenge after collecting a dataset is usually to obtain a set of phases, since only the structure factor amplitudes can be determined directly from the recorded diffraction images. Several methods exist, but when a structure of a closely related homologue is available, the most straightforward method is usually molecular replacement. With this method, initial phases are obtained by finding the best rotation and translation of the molecular model of the known structure in the unit cell of the unknown structure [103]. The method works best when the structure used as a search model is very similar to the unknown structure. While the best way to quantify the similarity between two structures would be root-mean-square deviation (RMSD) of the atomic positions, the only information that is available before the molecular replacement is the sequence identity. Typically, 30% sequence identity is considered the minimum for a good search model, although search models with a higher similarity might fail if the structures are too different and search models with lower similarity might succeed if the structures are similar [104].

For solving the structure of 33SSP–NB33_1, several potential search models were available. Structures had previously been published of two bestrophin homologues, namely chicken BEST1 (cBEST1) [37] and a prokaryotic bestrophin homologue from *Klebsiella pneumoniae* (kpBEST) [67]. Additionally, many nanobody structures were readily available in the Protein Data Bank. However, as the molecular mass of a nanobody is small (12.6 kDa), one nanobody was only expected to make up 1.3% of the total protein mass (948 kDa) in the asymmetric unit. Therefore the scattering contributed by a single nanobody would be similarly low, meaning that it would be very difficult to correctly place a nanobody using molecular replacement [105]. By comparison a 33SSP pentamer (174 kDa without bound nanobodies) was expected to contribute 18.3% of the total protein mass, meaning that it should ideally be easier to place correctly during the molecular replacement search. kpBEST was considered the best available search model, since it showed a higher similarity to 33SSP (approximately 25% sequence identity) than cBEST1 (sequence identity <20%).

In order to improve the search model, an alignment was produced between 33SSP and kpBEST. Based on this alignment, the side chains of non-conserved amino acid residues were truncated to the Cβ or Cγ position using the program SCULPTOR [106]. An initial molecular replacement search with PHASER [105] running on a single processor core did not return a solution within a few days and was therefore aborted. After testing various parameters, the

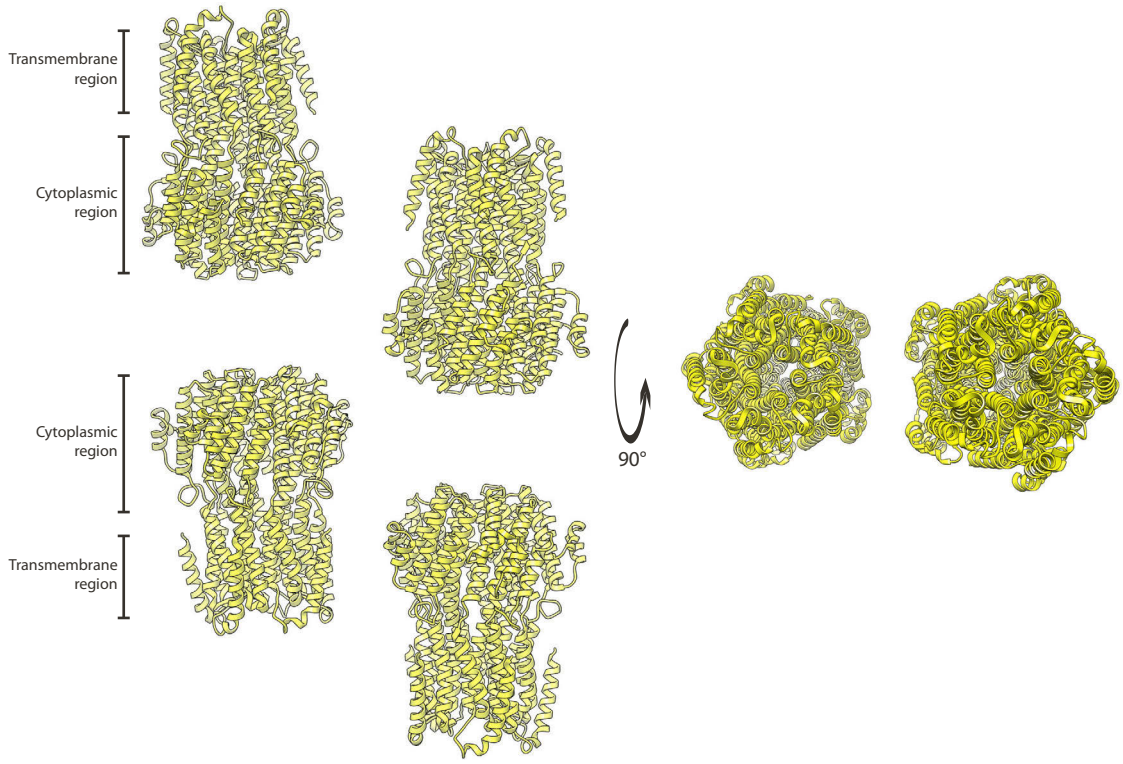


Figure 31 – Molecular replacement solution obtained with Phaser. The four pentamers present in the asymmetric unit are shown as yellow ribbons. For two of the pentamers, the cytoplasmic and transmembrane regions are labelled. A 90° rotation illustrates how one pentamer is placed directly above another, with the cytoplasmic regions facing each other.

expected RMSD between the search model and the unknown structure was lowered to 1 Å. Additionally, it was specified that placed components did not have to pack into a compact assembly, since the presence of detergent micelles and nanobodies could mean that very few crystal contacts would be present between the individual pentamers. After a lengthy search of 41 hours total processing time (the actual time was decreased by running on multiple processor cores), a solution was found.

In this solution, four pentamers were successfully placed in the asymmetric unit (Figure 31). Two pairs of pentamers were observed. In each pair, the cytoplasmic regions of two pentamers were facing each other, separated by approximately 30 Å. Within each pair, the axes of the pores were found to line up, so that one straight line could be drawn running down the middle of the pore in both pentamers. If such lines were drawn through both pairs, a slight tilt was observed, meaning that the two pairs were not completely parallel.

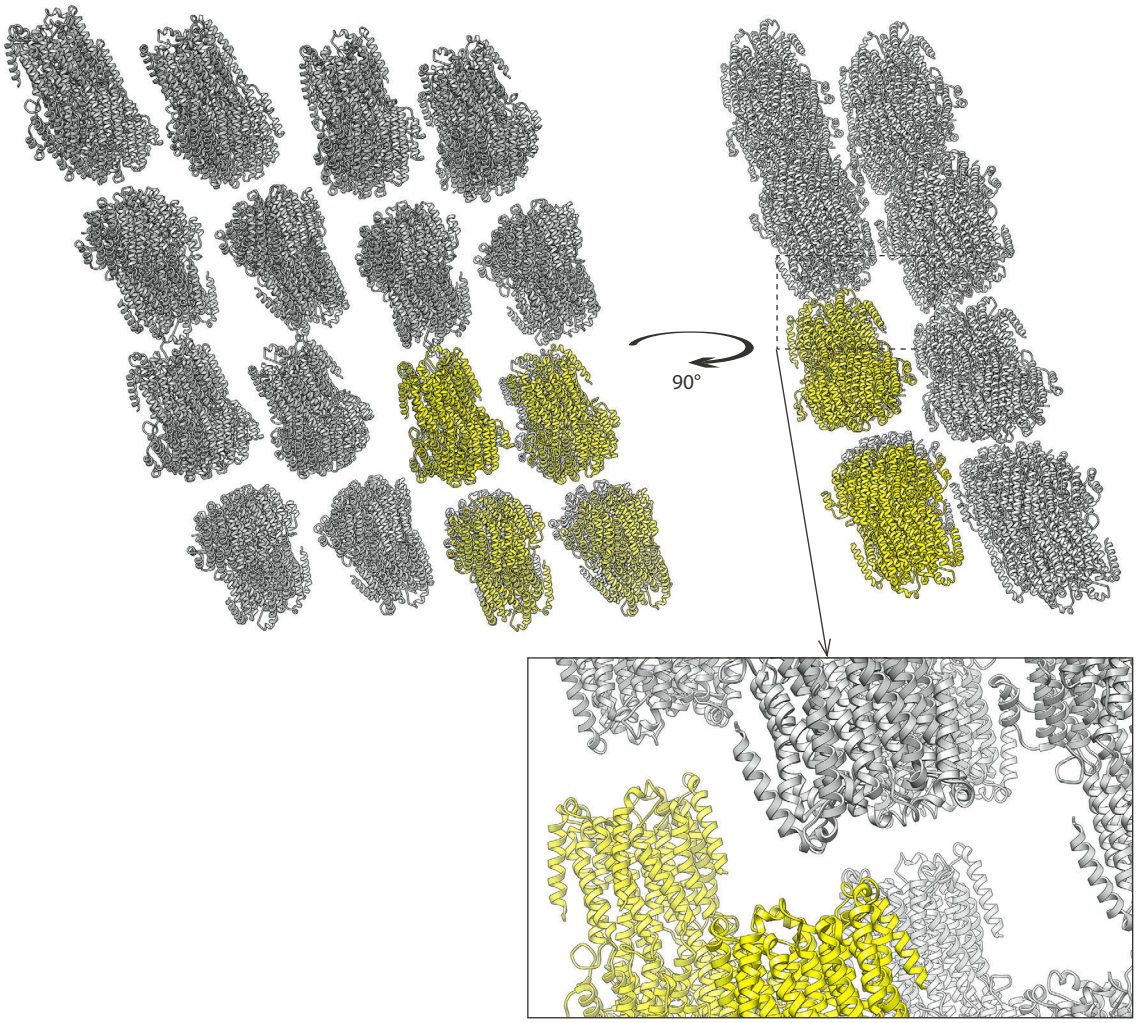


Figure 32 – Crystal packing of the molecular replacement solution obtained with Phaser. One asymmetric unit (the content of one unit cell) is shown in yellow. Seven additional asymmetric units are shown in grey. No potential crystal contacts could be observed.

Further analysing the crystal packing (Figure 32) did not reveal any potential crystal contacts between any of the pentamers. Enough space was observed around the transmembrane re-

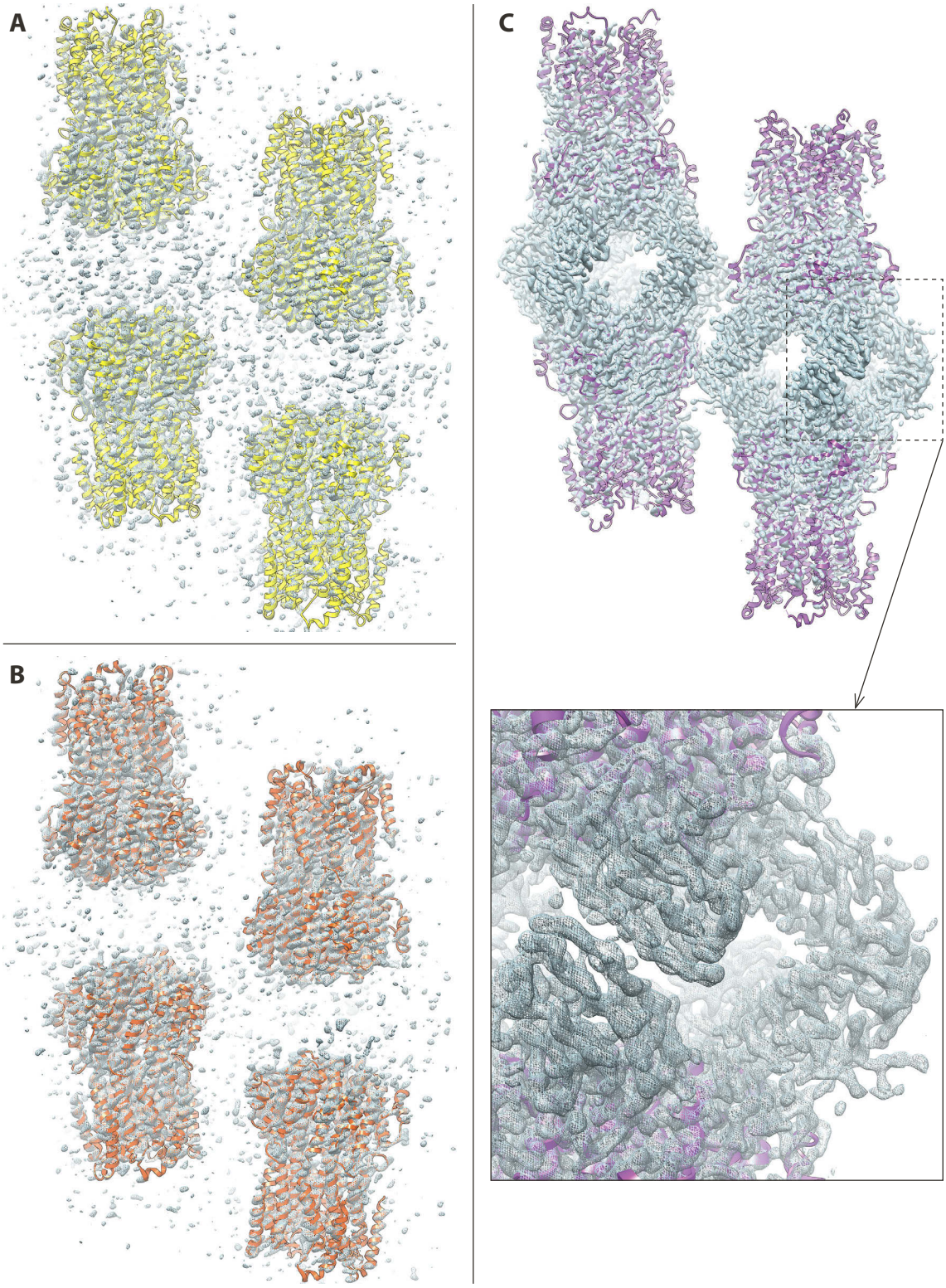


Figure 33 – Electron density maps and structures after molecular replacement and additional refinement or morphing. All maps (2mFo–DFc) are contoured at 2σ and shown as meshes with a light blue colour. The same region of the map, surrounding the bestrophin pentamers, is shown in each subpanel. Coloured ribbons represent the structures. **(A)** Solution obtained with Phaser. **(B)** Refinement of solution obtained with Phaser. **(C)** Morphing of solution obtained with Phaser.

gions to accommodate detergent micelles, as is typical for the so-called Type II membrane protein crystals [78]. For other proteins, the high concentration of lipid and detergent used during the crystallisation have sometimes been found to result in the formation of Type I crystals [82], where the packing resembles stacked 2D crystals, with the transmembrane regions packed much closer together. This did, however, not seem to be the case for the 33SSP–NB33_1 complex.

Since it was not known if the molecular replacement solution was correct, i.e. if the placement of the search model corresponded to the actual structure, the electron density map was examined closely. Density was seen around the helices (Figure 33a), but this was not surprising since the phases were expected to be biased towards the search model [107]. Surrounding the pentamers, the map was found to be rather noisy and no indication of nanobody binding could be observed.

As the next step, refinement of the solution was attempted. 5 cycles of restrained refinement in PHENIX [108,109] using NCS and secondary structure restraints, lowered the R-factors from 53%/53% (R-work/R-free) to 49%/52%. The electron density map did appear a little bit less noisy, but did still not reveal any new features (Figure 33b). Thus, it was still not clear if the solution was correct.

Finally an attempt was made to improve the model using iterated local density-guided model deformation and refinement, also known as morphing [110]. The actual morphing was performed automatically in the PHENIX software suite. In each cycle, a prime-and-switch map was calculated, since that type of map has been reported to be less affected by model bias than commonly used map types [111]. The model was then deformed to better fit the density and finally refined. This process was repeated 6 times. This procedure significantly improved the R-factors from 53%/53% (R-work/R-free) to 42%/46%. When inspecting the electron density map, a large improvement was observed (Figure 33c). In addition to the maps being less noisy, clear density for a total of 20 nanobodies could now be observed. 5

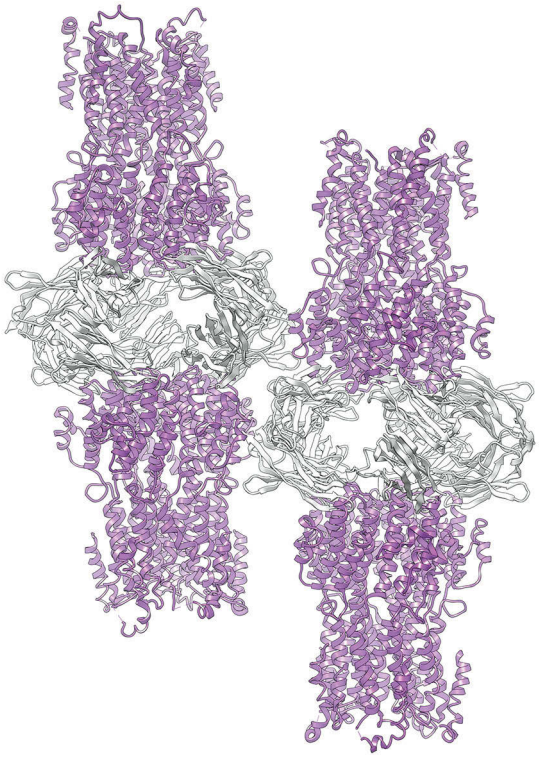


Figure 34 – Placement of NB33_1. The figure shows the output from the phased molecular replacement procedure used to place twenty nanobodies. The four 33SSP pentamers are shown as purple ribbons, while the twenty NB33_1 nanobodies are shown as white ribbons.

nanobodies appeared to be bound to the cytoplasmic region of each pentamer, and two pentamers appeared to be held together in a dimer of pentamers, due to nanobody–nanobody interactions.

To further improve the structure, a nanobody model was added in twenty copies. While it might have been possible to manually dock nanobodies into the density, an easier solution was chosen. Phased molecular replacement done using the program MOLREP [112], easily placed twenty nanobodies in the expected positions (Figure 34). The placed nanobodies all appeared to bind identically to 33SSP and the nanobody–nanobody interactions also appeared identical for each of the nanobody dimers.

2.4.3. Refinement

After placing all nanobodies correctly, the model was improved by iterative model building and refinement. Initial building efforts focussed on eliminating large errors. A loop on the extracellular side, between the first and the second transmembrane helix, was removed from each of the monomers in the structure as it was not supported by the density. This was not entirely surprising since this loop is four amino acid residues longer in 33SSP compared to kpBEST. Another surface loop between the third and the fourth transmembrane helices was also removed, in addition to a surface loop on the cytoplasmic side of the protein. Some of the side chains that had been removed in preparation for molecular replacement, were added back if supported by the density. For the nanobodies, the CDRs were found to be in poor agreement with the density and they were therefore also removed. Whenever possible, large segments of structure were copied using NCS operators, to avoid building the same parts twenty times.

Refinement was carried out in PHENIX using torsion-angle NCS restraints [113]. Initially, only coordinates and isotropic B-factors were refined. Refinement with grouped B-factors did not improve the refinement results, as judged by the R-factors. In later refinements, TLS parameters were also added. One TLS group was used for each chain, since adding more were not expected to lead to significant improvements, based on analysis of a refined model with the TLS-MD server [114].

At the time of writing, model building and refinement has still not been completed.

Resolution (Å):	47.92 - 3.1 (3.211 - 3.1)
Total reflections:	870561 (84178)
Unique reflections:	246398 (24632)
Used in refinement:	246218 (24622)
Used for R_{free}:	12314 (1232)
Wilson B-factor (Å^2):	86.97
R_{work}:	19.79 (35.37)
R_{free}:	23.06 (39.90)
No. of atoms:	62854
Protein residues:	8006
RMS deviations	
Bonds (°):	0.004
Angles (Å):	0.71
Ramachandran	
Favored (%):	97.80
Allowed (%):	1.97
Outliers (%):	0.23
Rotamer outliers (%):	3.75
Clashscore:	7.10
Average B-factor (Å^2):	104.81

Table 13 – Refinement statistics. Numbers in parentheses indicate the highest resolution shell and its statistics. 5% of the unique reflections were used for the R_{free} calculation.

The current refinement statistics are shown in table 13. R-factors are already quite low (R_{work} 19.79%, R_{free} 23.06%), possibly partially due to the extensive NCS restraints. The model still contains a higher than expected number of rotamer outliers, and certain surface loops are still missing from some of the pentamers. The number of amino acids placed in each subunit varies between the four pentamers present in the asymmetric unit. In the most complete pentamer all loops have been built, except for the surface loop connecting the first and second transmembrane helices. An initial trace of this loop, which is on the extracellular side of the protein, has been made for one of the subunits in this pentamer. The electron density for this loop is still weak, so the exact conformation of the loop and the conformations of the individual amino acid side chains are still not certain. Some electron density is, however, seen for some of the missing loops, so it is possible that these can eventually be built (Figure 35).

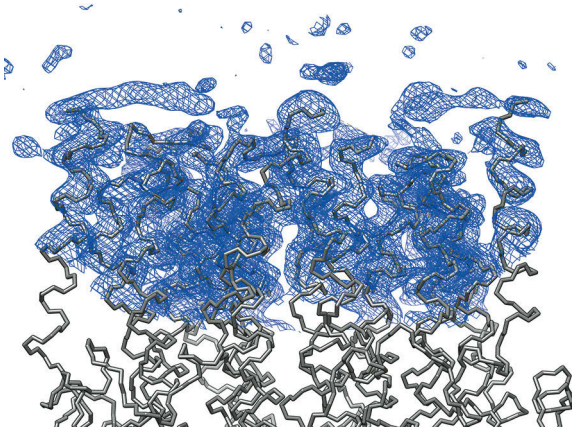


Figure 35 – Electron density of missing loops on the extracellular side of 33SSP. Electron density map (2mFo–DFc) contoured at 1σ and shown as a blue mesh. The structure of 33SSP is shown as a backbone trace, coloured in dark grey.

All nanobodies have been fully modelled, except for 5 unstructured amino acid residues on the N-terminus. Figures in the following sections have been prepared using the structure of the best modelled pentamer and its associated nanobodies.

2.5. Structure analysis

2.5.1. Overview of the 33SSP structure

As expected from the structures of other bestrophins, 33SSP is a pentameric, α-helical protein. The five monomers are arranged symmetrically around a central pore axis (Figure 36).

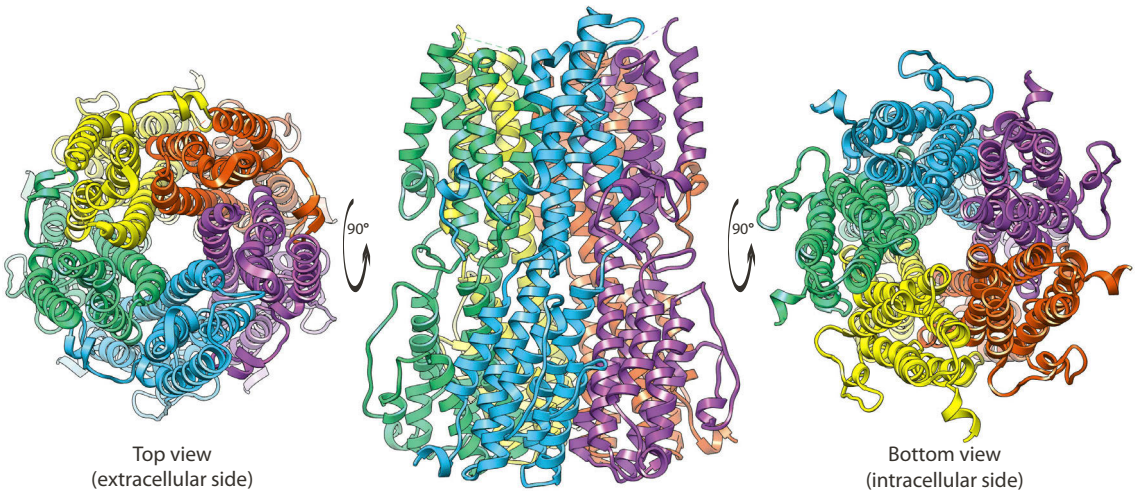


Figure 36 – Ribbon model of 33SSP. The figure shows the pentameric structure of 33SSP, with each monomer coloured in different colours. The middle part of the figure shows the protein viewed from the side, with the extracellular side on top and the intracellular side on the bottom. The left part of the figure shows the view from the extracellular side of the protein (top view), while right part shows the view from the intracellular side (bottom view). In the top and bottom views the pore is visible as a hole in the middle of the protein.

This can be observed as a hole in the middle of the protein when viewed from the top or bottom. Based on the understanding of animal bestrophin topology [115], the N-terminus and C-terminus are both expected to be located on the intracellular side. As certain helices are more hydrophobic than others, the membrane spanning part of the structure was estimated to correspond to the top of the protein, when viewed from the side as in figure 36. A more precise estimate of the position and orientation in the membrane was produced with the PPM web server [116], using a theoretical, complete model of 33SSP where the surface loop missing from four of the monomers had been copied from the complete monomer. The boundaries of the hydrocarbon core of the bilayer are shown in figure 37a. The thickness of the bilayer was estimated to be 27.2 ± 1.0 Å. By comparison the size of a complete pentamer is approximately 95 Å when measured along the pore axis and 75 Å across the cytoplasmic part of the protein. As illustrated in figure 37b, the protein consists of 8 longer α-helices. 4 of these helices are crossing the membrane and are labelled TM1–4. The first two and the last two of the transmembrane helices are separated by longer sequence regions constituting a folded intracellular component (Figure 37c). At the primary sequence level, the protein can thus not be divided into clearly separate domains. When looking at the structure, there are also no clear domain

boundaries since the second helix, which is lining the part of the pore located in the membrane, is connecting the transmembrane and cytoplasmic regions of the protein.

The N-terminus of each monomer is positioned close to the membrane interface and is wrapping around the protein (clockwise when viewed from the extracellular side) to the adjacent monomer, where a salt bridge is formed between Asp20 of one monomer and Lys222 of the

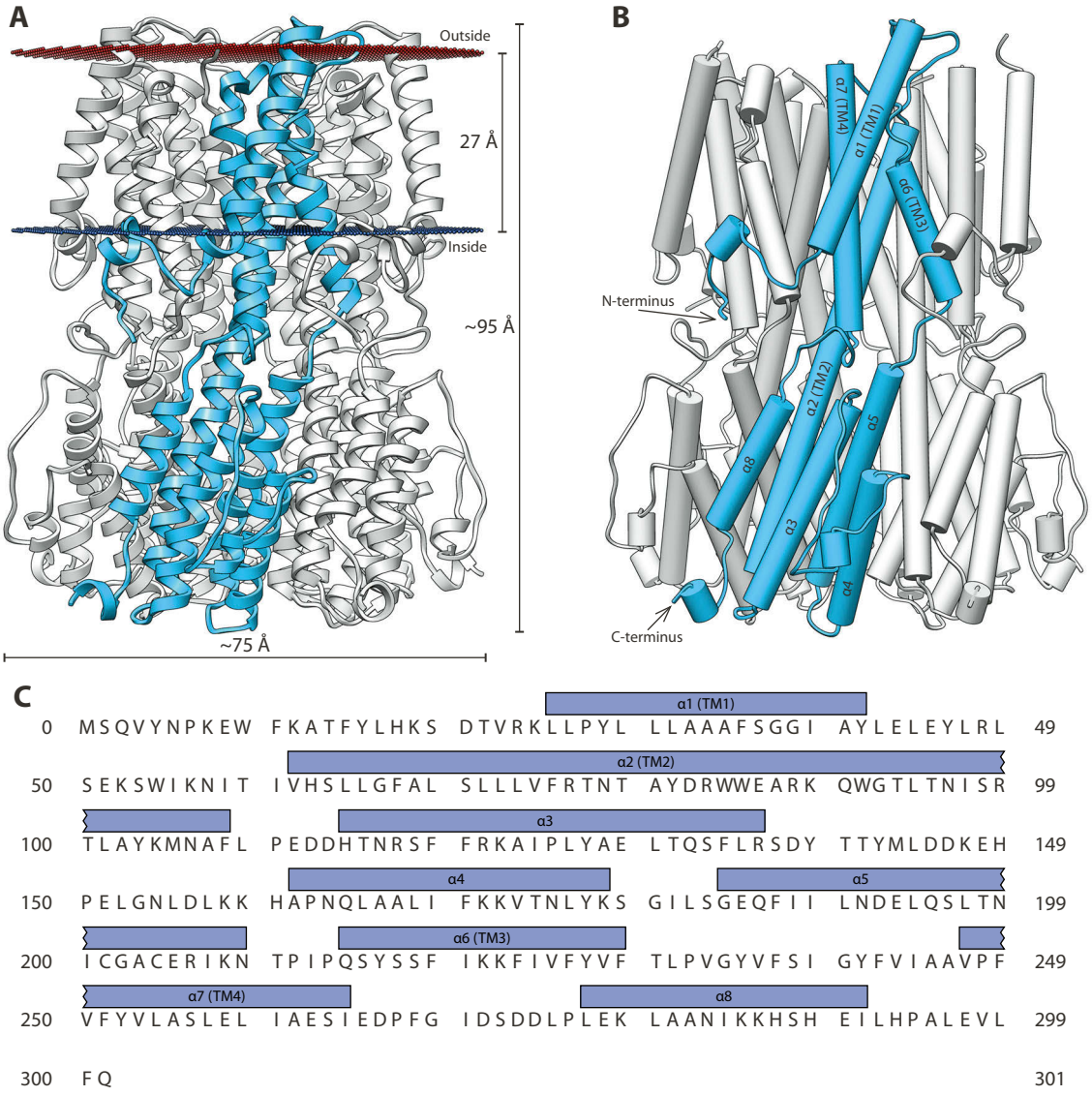


Figure 37 – Protein topology and sequence. **(A)** Based on the most complete monomer (highlighted in blue), a theoretical, complete model was generated and its position in the membrane was estimated. Sheets of red and blue balls indicate the boundaries of the hydrocarbon core of the bilayer. The extracellular side is labelled "outside" and the intracellular side "inside". The approximate thickness of the bilayer is indicated as well as the approximate dimensions of the pentamer. **(B)** Structure of 33SSP with helices shown as cylinders. One monomer is highlighted in blue and helices longer than five amino acid residues are labelled. Positions of the N-terminus and C-terminus are indicated. **(C)** Sequence of the 33SSP construct used for crystallisation. Helices are annotated and labelled in accordance with the model in the previous panel. The second amino acid residue on the N-terminus and the last six amino acid residues on the C-terminus were introduced by cloning. The numbering was started at zero to account for the N-terminal addition. The first three amino acid residues on the N-terminus and the last amino acid residue on the C-terminus are not modelled in the structure. The loop between the first and the second transmembrane helices is only fully modelled in one of the monomers (highlighted in blue).

adjacent monomer. Additional intermolecular salt bridges are observed from Glu111 to Lys171 and from Arg99 to Glu205, all in the intracellular region of the protein. Furthermore, the pentameric complex is stabilised by an average of 22 intermolecular hydrogen bonds. The interface area between two monomers measures approximately 3000 Å^2, as calculated by the PDBePISA webserver [117].

Some parts of the structure appear to be more flexible than others. When analysing the crystallographic B-factors, these were found to be significantly higher in the transmembrane region compared to the core of the cytoplasmic region. This is illustrated in figure 38 where the average B-factors for each residue are mapped on a cartoon model of 33SSP. In the cytoplasmic region, the B-factors are generally low, except for loops on the surface. Getting closer to the membrane the B-factors appear to increase. The residues close to the outside and the extracellular loops, consistently appear to have B-factors much higher than average, indicating a higher degree of flexibility than the rest of the protein. Consistent with this observation, the structure has been more difficult to build closer to the extracellular side, where the electron density maps appear less well-defined, compared to the cytoplasmic region.

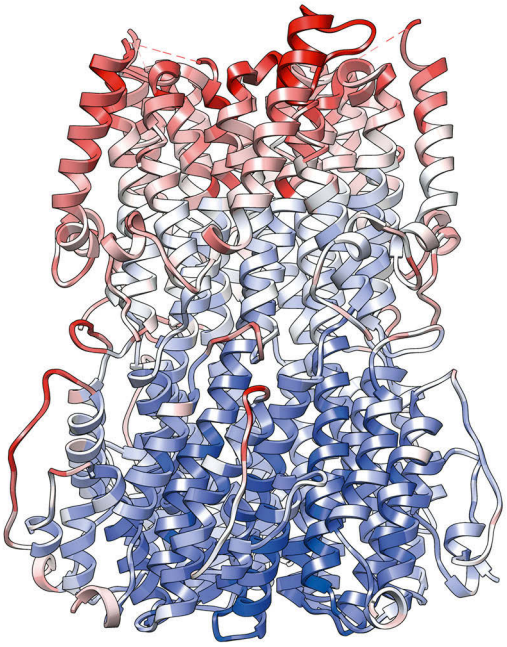


Figure 38 – Distribution of B-factors in the 33SSP structure. B-factors were converted to isotropic B-factors, since refinement had been performed using TLS parameters. The average B-factor for each amino acid residue is displayed on the cartoon model of 33SSP, using colours ranging from blue (40 Å^2) over white (105 Å^2) to red (170 Å^2). The average B-factor for the entire 33SSP–NB33_1 structure was calculated as 105 Å^2.

Overall, the structure of 33SSP looks similar to the known cBEST1 and kpBEST structures. The topology is analogous and when positions on the surface of the proteins are excluded, the core helices are found to be conserved between all three homologues. In an attempt to compare the structures directly, the pentameric cBEST1 and kpBEST structures were each superimposed on the pentameric 33SSP structure using secondary structure matching [118] in Coot [119]. In this case the whole pentameric structures were rotated and translated as rigid bodies to provide the best fit to the 33SSP structure. The root-mean-square deviation of the Cα positions (core RMSD) was 3.76 Å for the superimposition of cBEST1 on 33SSP and 2.96 Å for kpBEST on 33SSP. These rather large deviations were also obvious upon visual inspection.

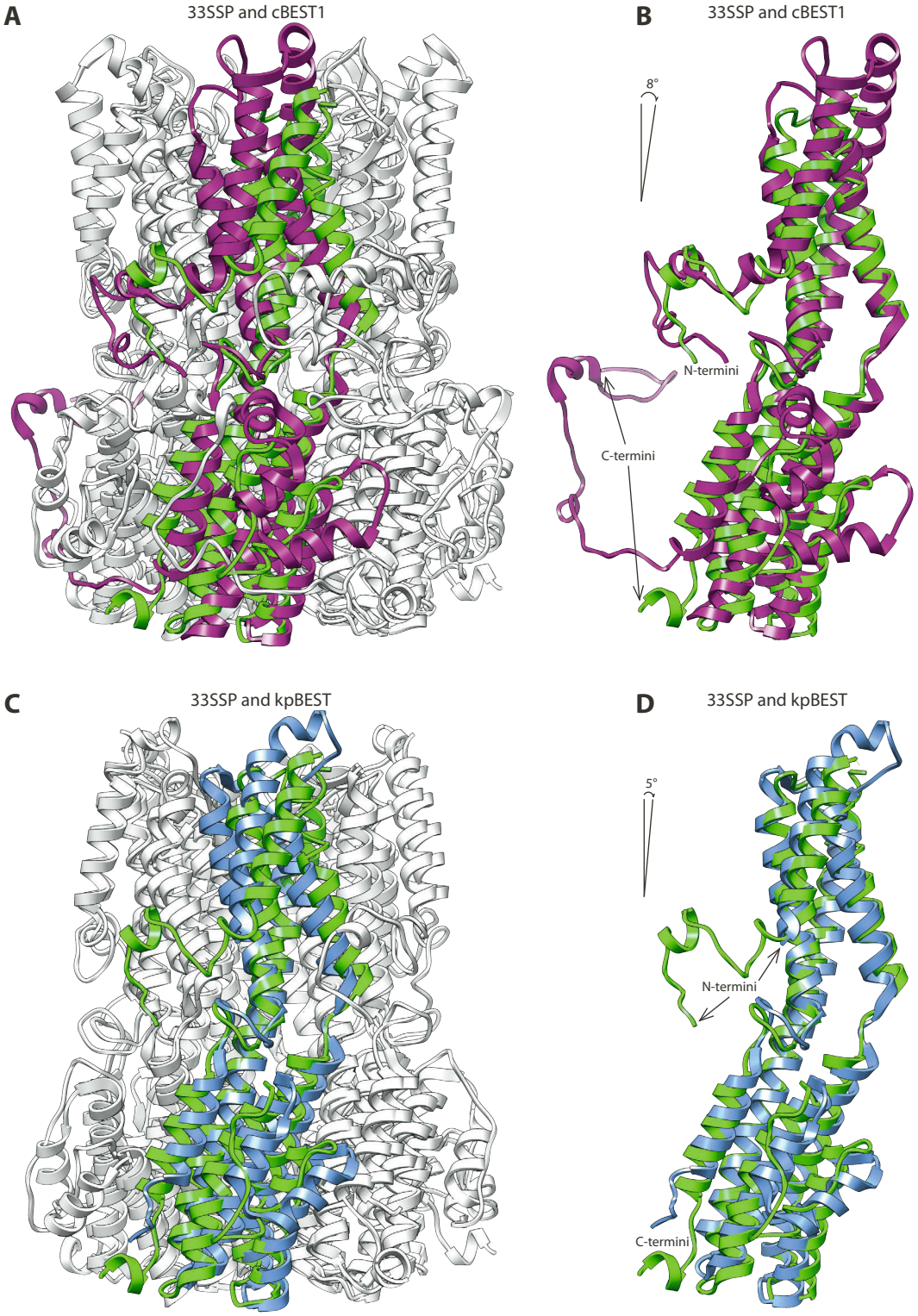


Figure 39 – Structures of cBEST1 and kpBEST superimposed on the structure of 33SSP. Structures were superimposed using secondary structure matching (SSM) in Coot. For the highlighted monomers, green is used for 33SSP, dark purple for cBEST1 and blue for kpBEST1. Furthermore, the positions of the N- and C-termini are indicated. **(A)** A cBEST1 pentamer superimposed on a 33SSP pentamer. One monomer of each structure is highlighted. **(B)** A monomer of cBEST1 superimposed on a 33SSP monomer. The approximate rotation of the cBEST1 monomer compared to the previous panel is indicated. **(C)** A kpBEST pentamer superimposed on a 33SSP pentamer. One monomer of each structure is highlighted. (**D)** A monomer of kpBEST superimposed on a 33SSP monomer. The approximate rotation of the kpBEST monomer compared to the previous panel is indicated.

For both cBEST1 (Figure 39a) and kpBEST (Figure 39c) the structures were similar enough for a credible superimposition, but significant shifts were observed for most helices. However, when a monomer of either cBEST1 (Figure 39b) or kpBEST (Figure 39d) was superimposed on a monomer of 33SSP, the structures were found to align much better. In this case, the core RMSD values were 2.83 Å or 2.08 Å for the superimposition of cBEST1 or kpBEST, respectively. When comparing the position of a cBEST1 monomer that had been superimposed as part of a pentamer, with a monomer that had been superimposed as a monomer, the difference was found to be a rotation of approximately 8°, around a point located between the cytoplasmic and transmembrane regions. For kpBEST a similar, but smaller, rotation of 5° was observed. It thus seems, that while the overall pentameric assemblies are similar, the exact orientation of the monomers with respect to each other, can vary between different homologues.

When comparing the structure of 33SSP to the structure of kpBEST (Figure 39d), a striking difference was found in the N-terminus. Upon comparison of the amino acid sequences, the two homologues were found to have N-termini of similar lengths. However, in the structure of kpBEST, the N-terminal region preceding the first transmembrane helix does not appear to be structured. By comparison, 33SSP and cBEST1 seem to have N-termini of similar lengths, both located close to the membrane (Figure 39b). A discussed later, the N-terminus is involved in calcium binding in cBEST1. Additionally, cBEST1 has a significantly longer C-terminus, which wraps around the pentamer, covering two adjacent monomers.

2.5.2. Ion conduction pathway

As previously discussed, the pore of 33SSP is straightforward to identify as a hole in the middle of the protein. Along the axis of this pore, narrow constrictions can be observed in two separate places. Towards the extracellular region of the protein, a constriction is formed by three residues in the second transmembrane helix. Looking along the pore axis from the outside of the protein, a constriction consisting of Phe67, Ile71 and Phe75 is clearly visible. Due to the fivefold symmetry of the protein, each of these residues are observed five times at the pore, where they are arranged symmetrically around the pore axis (Figure 40a). This region of the protein has previously been referred to as the "neck", while another constriction has been referred to as the "aperture" [37]. Both terms will be used in the following when referring to these constrictions. The aperture can be observed when viewing the protein from the intracellular side. Here the residues Glu185 and Ile188 form a constriction, again with fivefold symmetry around the pore axis (Figure 40b). Whereas the neck is made up of hydrophobic residues in all three available bestrophin structures, the aperture of 33SSP is unique in having a ring of negatively charged residues at the position closest to the cytoplasm. At this position either a valine or isoleucine residue can be found in cBEST1 or kpBEST, respectively. Also unlike the other bestrophin structures, the aperture in 33SSP is made up of two residues. At the

position of Ile188, cBEST1 and kpBEST feature either a glutamine residue that does not point toward the pore axis or a glycine residue, respectively.

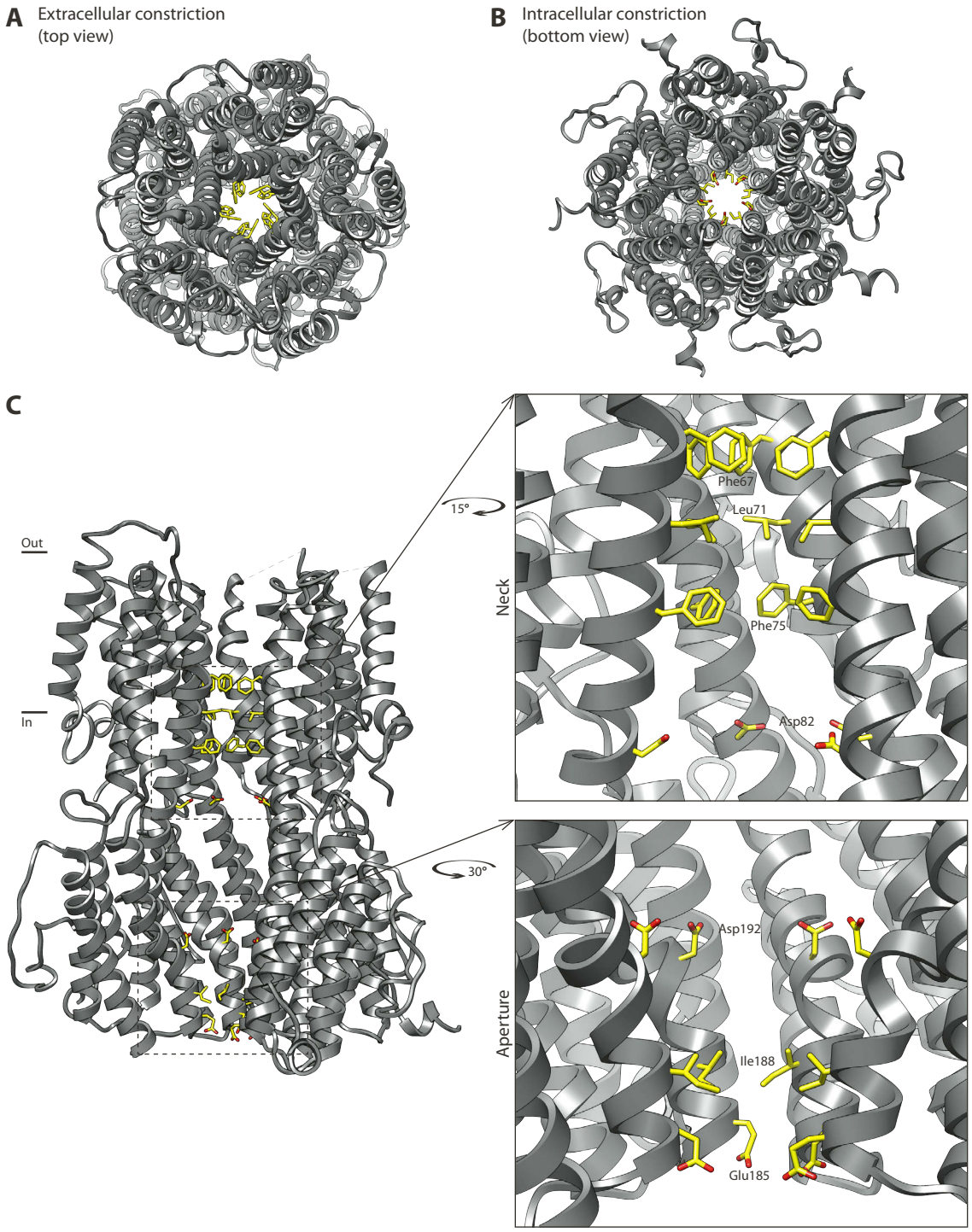


Figure 40 – View along the pore of 33SSP. Structure of 33SSP shown as grey ribbons. Highlighted residues are shown as yellow sticks. **(A)** Extracellular view of 33SSP (top view). Phe67, Leu71 and Phe75 are shown as yellow sticks. **(B)** Intracellular view of 33SSP (bottom view). Glu185 and Ile188 are shown as yellow sticks. **(C)** View of 33SSP from the side. Approximate membrane boundaries are shown. One monomer and the N-terminus of an adjacent monomer are hidden, resulting in a free view to the pore. From the top to the bottom the following residues are shown as yellow sticks: Phe67, Leu71, Phe75, Asp82, Asp192, Ile188 and Glu185. The neck and aperture regions are rotated and enlarged for a clearer view.

Between the neck and the aperture, the structure contains a large inner cavity. Interestingly, in 33SSP two rings of negative charges can be found in this inner cavity. One ring is made up of Asp82 and located close to the neck. The other ring is made up of Asp192 and located close to the aperture (Figure 40c). These rings of negative charge together with the negative charges at the aperture contribute to a negative surface potential in the inner cavity of 33SSP, as described later.

Most of the pore lining residues in the neck and aperture regions are well defined in the electron density as illustrated in figure 41. Only for the outermost residue (Phe67), the density is not clearly defining its exact conformation. Some density is, however, visible and other conformations consistent with that density would probably lead to clashes due to the narrow constriction at this position.

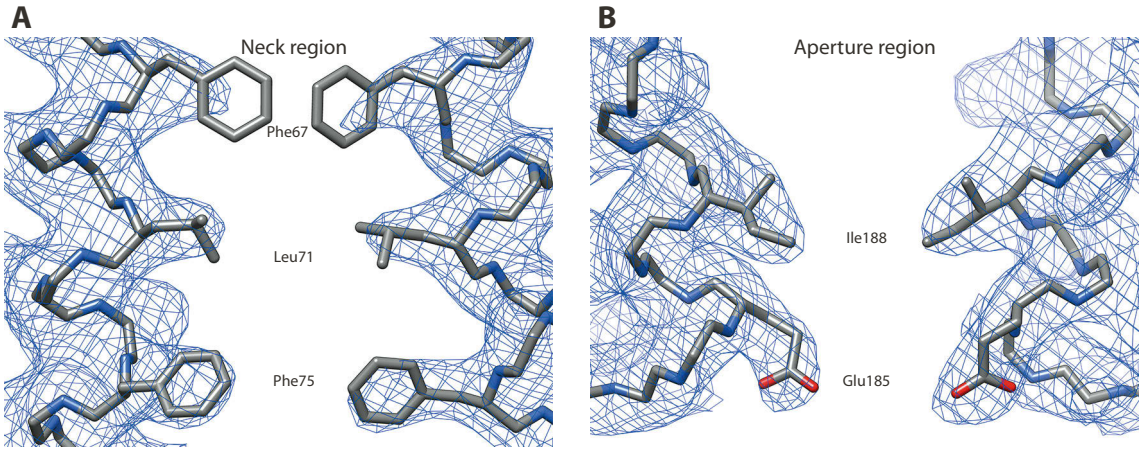


Figure 41 – Electron density of pore constrictions in the structure of 33SSP. Electron density map (2mFo–DFc) contoured at 1σ and shown as a blue mesh. The structures of two 33SSP monomers are shown as backbone traces, coloured in dark grey. Selected side chains are shown and labelled. **(A)** Constriction in the membrane region (neck region). **(B)** Constriction on the intracellular exit from the pore (aperture region).

To further analyse the constrictions at the neck and aperture, the pore dimensions of 33SSP as well as cBEST1 and kpBEST were calculated using the program HOLE [120]. The program measures the largest sphere that can fit at various positions sampled along the pore axis. The results presented in figure 42a show that the tightest constriction in 33SSP is found at the outermost position of the neck, where Phe67 is constricting the pore to a radius of 1.05 Å. Leu71 (middle position) and Phe75 (innermost position) lead to constrictions of 1.62 Å and 1.25 Å, respectively. In the aperture region, 33SSP is found to be less constricted with the radii measuring 1.58 Å at Ile188 and 1.59 Å at Glu185.

By comparison, the neck region of cBEST1 (Figure 42b) is constricted to a similar degree at the middle position (Phe80) as 33SSP at the outermost position (Phe67). At the innermost position of the neck, cBEST1, however, is less constricted with a pore radius of 2.02 Å. In the

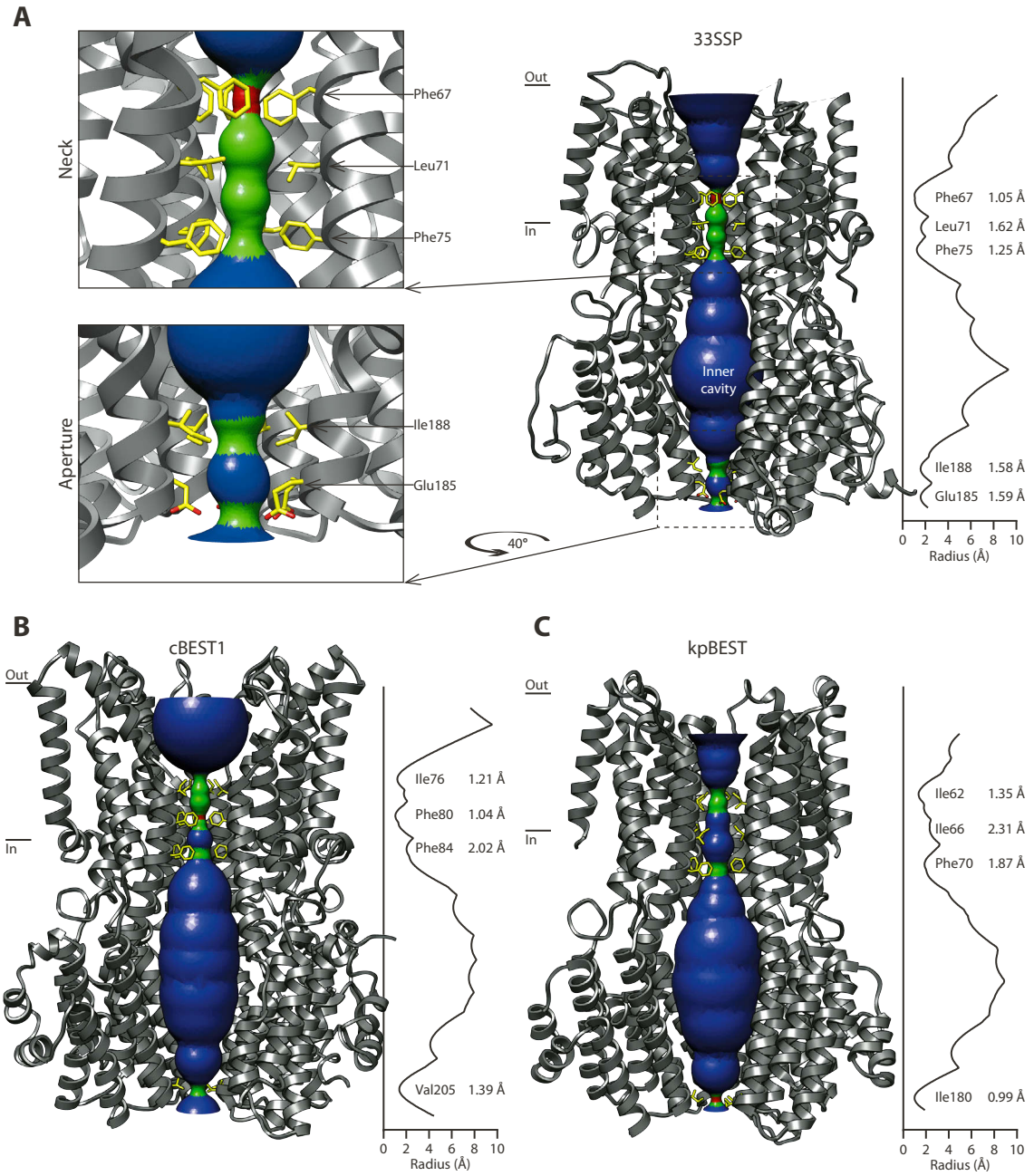


Figure 42 – Pore dimensions of 33SSP compared to other bestrophin structures. Pore dimensions were calculated using the program HOLE and illustrated as solid surfaces coloured red where the pore is too narrow for a water molecule (< 1.15 Å), green where a single water molecule can fit (1.15 Å > radius < 2.30 Å) and blue where more than one water molecule can fit (> 2.30 Å). The protein structures are shown as grey ribbons. Approximate membrane boundaries are shown. To aid in visualisation, one of the monomers and part of the N-terminal and C-terminal sequences from an adjacent monomer are hidden from each structure, resulting in free views to the pores. Residues located at the pore constrictions are shown as sticks and coloured yellow. For each structure, a plot showing the pore radius is aligned along the length of the pore axis. **(A)** Pore dimensions of 33SSP. The neck and aperture regions are enlarged. **(B)** Pore dimensions of cBEST1. **(C)** Pore dimensions of kpBEST.

aperture region, cBEST1 appears to be slightly more constricted than 33SSP, but only at one position, unlike the double constriction found in the aperture of 33SSP.

The neck region of kpBEST is the least constricted of the three compared channels (Figure 42c). Here the narrowest point is at the outermost position (Ile62) where the radius measures

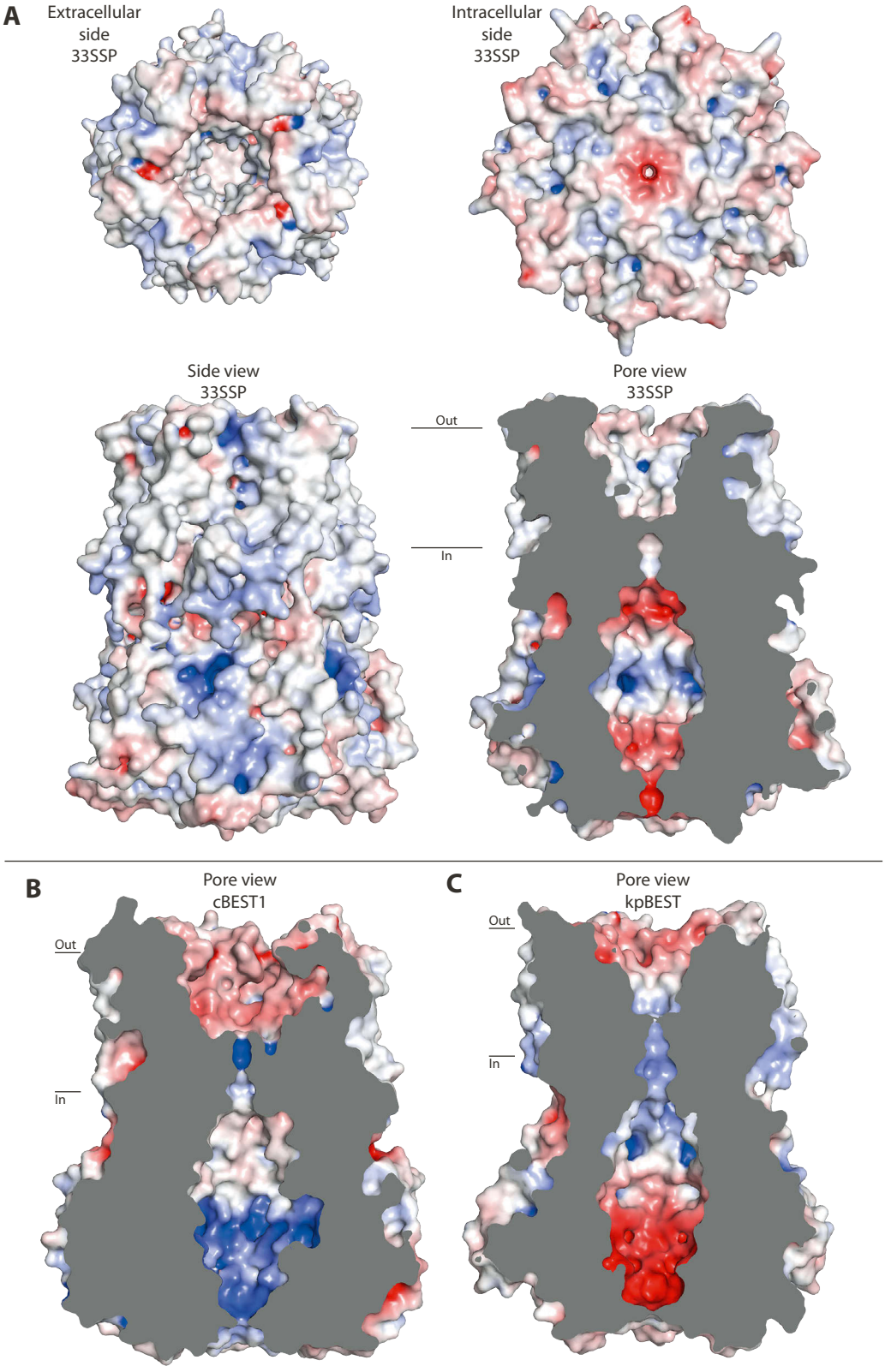


Figure 43 – Electrostatic surface plots of 33SSP compared to other bestrophin structures. For 33SSP, a theoretical, complete model, based on the most complete monomer, was generated and used for the calculation. Surface models of 33SSP as well as cBEST1 and kpBEST are coloured by electrostatic potential (red: –10 kT e^{-1}, white: neutral, blue: 10 kT e^{-1}). For the side views, the membrane positions are indicated. **(A)** 33SSP shown from the top (extracellular side), from the bottom (intracellular side) and from the side. Additionally, a view of the pore is shown, where half of the structure has been cut away. **(B)** Pore view of cBEST1. **(C)** Pore view of kpBEST.

1.35 Å. Contrary to the neck, kpBEST is found to be more constricted than either 33SSP or cBEST1 at the aperture where it measures 0.99 Å.

Comparing the pore lining residues in the necks of the three structures, it is evident that they all contain hydrophobic residues. The only position that appears to be fully conserved between the three channels is the innermost position, where a phenylalanine residue is found in each structure. This position also appears to be highly conserved when comparing to other bacterial bestrophins. Out of the 128 bacterial homologues that were selected for screening, 121 (95%) had a phenylalanine at this position. In the remaining 7 sequences, either isoleucine, leucine or tyrosine was found. The middle position appears less conserved, but typically contains a hydrophobic residue. 117 out of the 128 (92%) selected homologues contained either leucine, isoleucine, phenylalanine or valine at this position. The outermost position was also mostly hydrophobic with 86 of the 128 (67%) selected homologues containing phenylalanine, isoleucine, leucine or valine. However, 35 out of the 128 (27%) had either threonine or serine at this position. It thus appears to be the least conserved position.

To further investigate the charge distribution, the electrostatic surface potential was calculated for 33SSP as well as cBEST1 and kpBEST. On the surface of 33SSP patches of positive and negative potential can be observed, but around the pore entry no charged binding pockets are found (Figure 43a), contrary to what has been reported for cBEST1 [37]. Looking inside the pore, a strong negative potential can be observed at both ends of the inner cavity, close to the neck and aperture, respectively. By comparison cBEST1 has a strong positive potential (Figure 43b) and kpBEST a strong negative potential (Figure 43c) near the aperture. For 33SSP, the region inside the neck constriction also contains positive potential, but probably this is highly uncertain since this region is located close to the membrane interface and therefore could be expected to be strongly affected by the membrane electric field, which is not accounted for in these calculations. Therefore, to get a more detailed and accurate description of the electrostatics, more advanced calculations of the electrostatic potential would be required, taking the presence of a membrane and the associated electric field into account. Also, if the structure of 33SSP represents a closed channel, the electrostatics could change when the channel opens, if the opening involves large conformational changes.

2.5.3. Ligand binding loop and N-terminus

cBEST1 is activated by calcium like the mammalian bestrophins [37]. As previously described, calcium ions are bound by the so-called calcium-clasp, which includes a stretch of five acidic residues (Glu-Asp-Asp-Asp-Asp), and a part of the N-terminus. In 33SSP this sequence is only partially conserved (Ile-Asp-Ser-Asp-Asp) and since kpBEST has been reported not to be activated by calcium, it is uncertain if 33SSP would be. 33SSP has a loop in the same location and of the same length as the calcium-clasp in cBEST1, but it is uncertain if an activating li-

gand would bind in this location or somewhere else. It is nonetheless a possibility, so this loop will be referred to as the proposed ligand binding site.

The proposed ligand binding site appears in two different conformations in the structure of 33SSP. In one conformation, part of the loop is pushed in the direction of the last transmembrane helix. The possible binding pocket seems closed or collapsed in this conformation (Figure 44a). In the other conformation, the possible binding pocket appears more open (Figure 44b). In order for the proposed ligand binding site to adopt the open conformation, another loop on the surface of the cytoplasmic region, must move away. This loop, starting around position 140 and therefore referred to as the 140-loop, would otherwise partially clash with the ligand binding site (Figure 44c–d).

As the protein complex is crystallised with EGTA it is unlikely that calcium could be bound in the crystal. Since no other divalent ions or other potential ligands were added to the crystallisation condition either, the two conformations of this loop are most likely both ligand-free

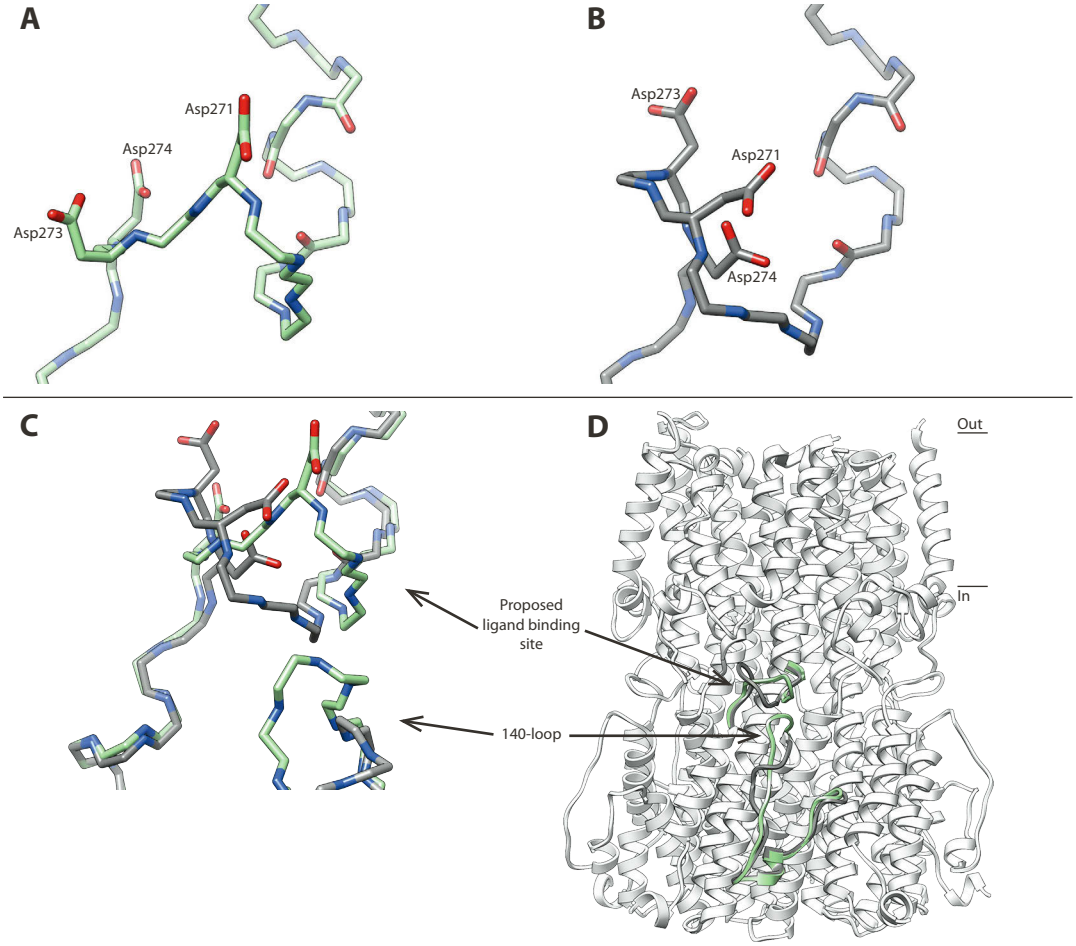


Figure 44 – Structure of the proposed ligand binding site of 33SSP. **(A)** Collapsed conformation of the proposed ligand binding site (light green). Selected side chains and main-carbonyls are shown. **(B)** Open conformation of the proposed ligand binding site (dark grey). Selected side chains and main-carbonyls are shown. **(C)** The two different conformations of the proposed ligand binding loop are superimposed. The 140-loop is shown to illustrate how this loop moves to accommodate the open conformation of the proposed ligand binding site. **(D)** Overview showing the proposed ligand binding site of 33SSP and the 140-loop in the two different conformations.

conformations. The loop thus appears to move between these two conformations in the absence of a ligand. It is possible that binding of a ligand would lock the loop in one of these two conformations or in another conformation.

The two different loop conformations appear well-defined by the electron density map, even though density is not clearly visible for all side chains (Figure 45). For the other monomers in the structure, the proposed ligand binding loop can be assigned to one of the two categories, although some small differences are observed. Also, in some monomers this particular loop appears less well defined in the electron density map, with density missing for most of the side chains.

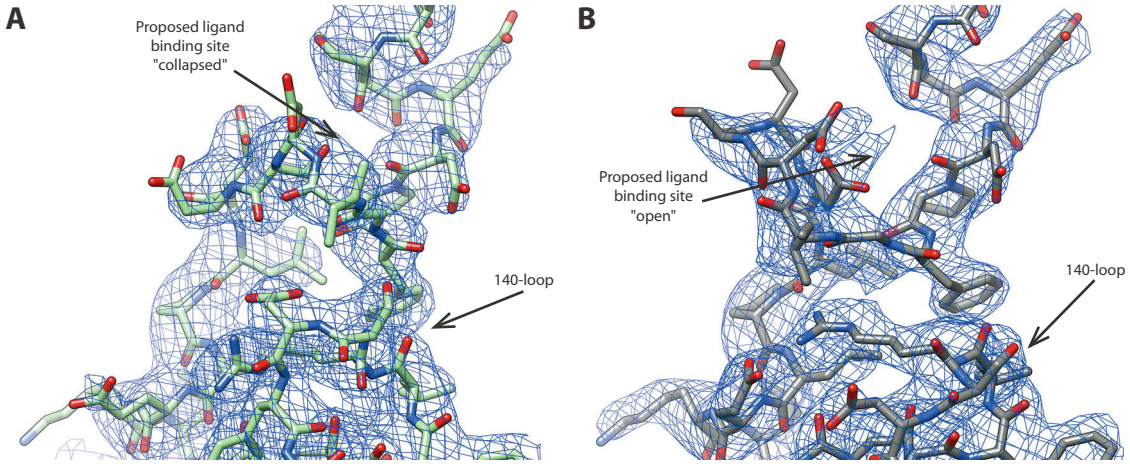


Figure 45 – Electron density of the proposed ligand binding site of 33SSP. Electron density map (2mFo–DFc) contoured at 1σ and shown as a blue mesh. The proposed binding sites of two different monomers showing different conformations are displayed. Additionally, the loops starting around residue number 140 are shown. **(A)** Proposed ligand binding site in the "collapsed" conformation. **(B)** Proposed ligand binding site in the "open" conformation.

2.5.4. Nanobody structure and binding

As previously described, 33SSP was crystallised in complex with a nanobody. The structure contains 20 nanobodies in the asymmetric unit. The structures of the individual nanobodies are largely identical, with only small variations observed. Typically, those variations are observed in regions that are involved in crystal packing. As expected from other nanobody structures [121,122], the structure consists of a sandwich of two β-sheets (Figure 46) that are connected by a disulphide bridge. The arrangement of the individual strands is consistent with a classical immunoglobulin fold [123], and the strands in figure 46 have thus been labelled using the immunoglobulin nomenclature. The only deviation from the described is found in strand A, which appears to be split in half (based on DSSP [124] secondary structure assignments) with each half of this strand being part of a different β-sheet. For this reason, the second half of strand A has been labelled A' in figure 46. This split of the first strand can also be observed when examining other nanobody structures (e.g. PDB accession numbers 1MEL [122]

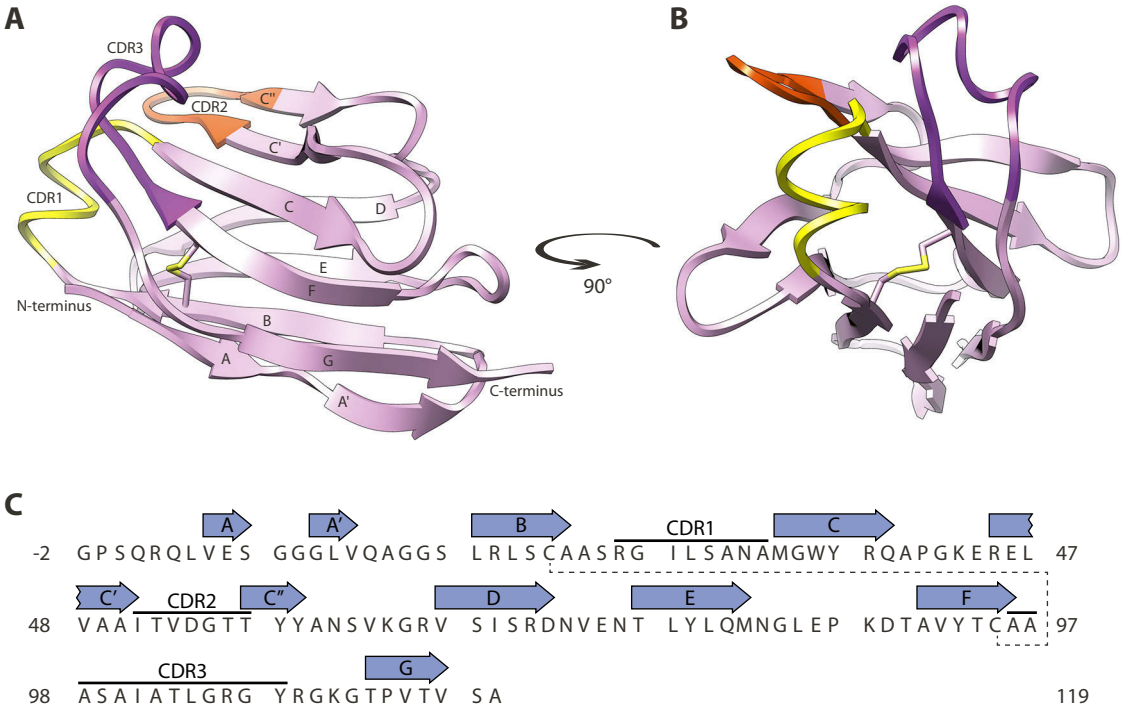


Figure 46 – Structure and sequence of one nanobody from the 33SSP–NB33_1 complex. **(A)** The nanobody is shown as a ribbon model. CDR1 is coloured yellow, CDR2 red and CDR3 purple. β-strands are labelled with the strand names typically used for immunoglobulin folds. Additionally, the name A' is introduced for the strand between strand A and strand B. The strands form two β-sheets, which pack against each other in a sandwich structure. The lowermost β-sheet is formed by the strands A, B, E and D, while the uppermost is formed by the strands A', G, F, C, C' and C''. A conserved disulphide bridge, linking the two sheets, is shown between strand B and strand F. **(B)** 90° rotation of the nanobody. **(C)** Sequence of the nanobody construct used for crystallisation. 3 first amino acid residues on the N-terminus and the 1 last amino acid residue on the C-terminus were introduced by cloning. The numbering was started at –2 to account for the N-terminal addition. 5 amino acid residues on the N-terminus are not modelled in the structure. β-strands are annotated as blue arrows and the positions of the CDRs are indicated with black lines. A dashed line shows the position of the disulphide bridge.

or 1HCV [121]) and is therefore not unusual. As expected the two β-sheets are anti-parallel, except for the second half of strand A which associates with strand G in a parallel orientation.

As expected the three CDR loops are located on one end of the nanobody, which binds to the target protein (33SSP). Each nanobody is binding on the cytoplasmic region of 33SSP (Figure 47a–b), with five nanobodies per pentamer that are tilted with respect to the pore axis. The epitope is remote from the pore and the putative ligand binding sites, and it thus does not seem likely that binding of the nanobody would affect the function of the channel. The binding occurs at the interface between two adjacent 33SSP monomers, where one nanobody interacts with both monomers (Figure 47c–d). For each nanobody, the interface with one monomer buries 750 Å^2 and with the other monomer 300 Å^2 of its surface (both calculated by the PDBePISA webserver).

As previously discussed, the nanobodies bring two pentamers in contact, thereby forming a complex, which can be described as a dimer of pentamers. This complex could be observed

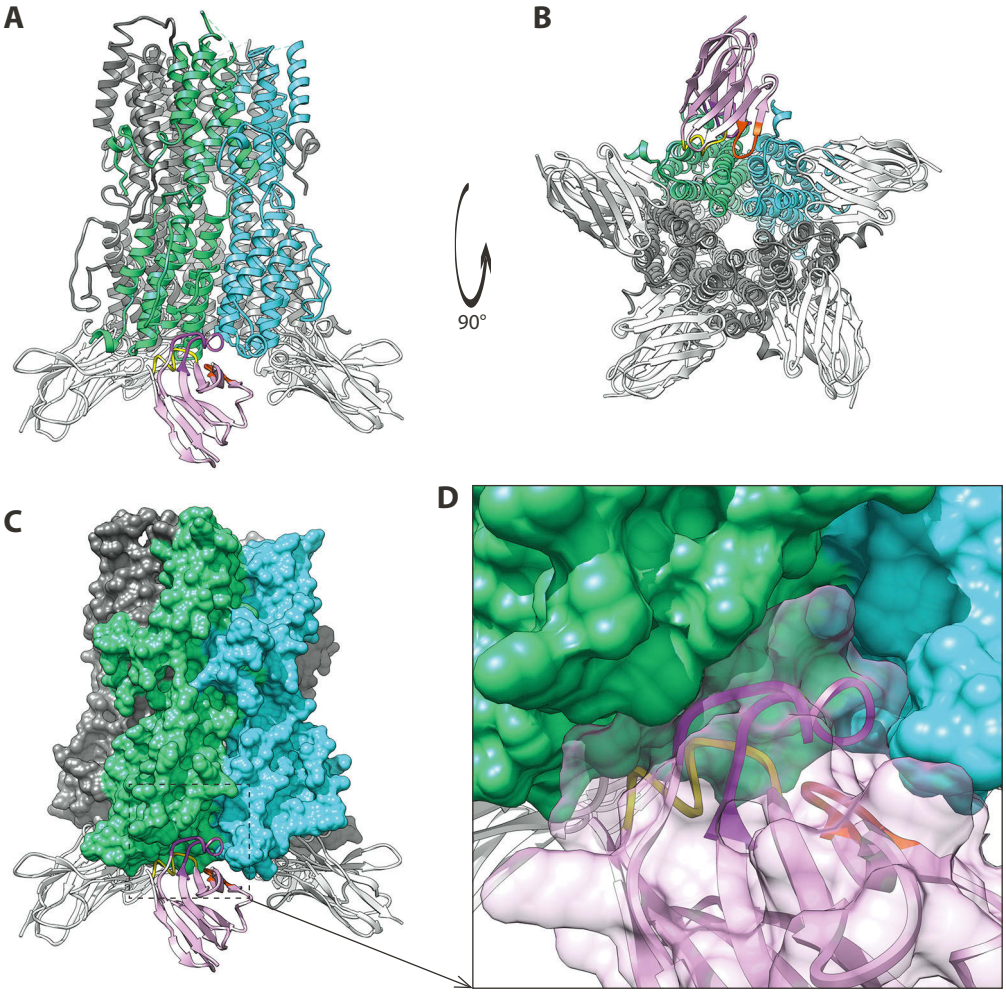


Figure 47 – Binding of NB33_1 to 33SSP. One pentamer shown in different representations with five nanobodies bound. The two 33SSP monomers closest to the front are coloured green and cyan; the remaining monomers are coloured grey. The nanobody closest to the front is coloured pink; the remaining nanobodies are coloured white. For the front nanobody, CDR1 is coloured yellow, CDR2 red and CDR3 purple. **(A)** Ribbon representation of the pentamer and the five nanobodies, seen from the side. **(B)** 90° rotation showing the pentamer from the cytoplasmic side. **(C)** Surface view of the pentamer with the nanobodies shown as ribbons, with the same orientation as in panel A. **(D)** Enlargement of the region indicated on the previous panel. The surface of the nanobody is shown as a transparent pink cloud.

during purification as a peak that eluted at low volumes from the gel filtration column. However, when NB33_1 was purified alone it eluted at similar volumes as the other tested nanobodies, indicating that it was monomeric. Even when injecting the nanobody at high concentrations (approximately 20 mg/ml), there were no indications that it was forming dimers. An examinination of the the nanobody–nanobody interface that is observed in the crystal structure did also not reveal anything that resembled a high-affinity binding surface. The two nanobodies are interacting through hydrogen bonds between residues that are mainly located in the first two strands (Figure 48). No salt bridges were observed and the interface area is about 400 Å^2 in size (calculated by the PDBePISA webserver). The observed dimer of pentamers is therefore likely formed due to an avidity effect, where the sum of five weak nanobody–nanobody interactions is enough to form a stable complex. The sequence of the interface region

appears to be identical to the other tested nanobodies, so the fact that the dimer of pentamers complex could form is probably largely due to the binding orientation of this particular nanobody on the 33SSP pentamer.

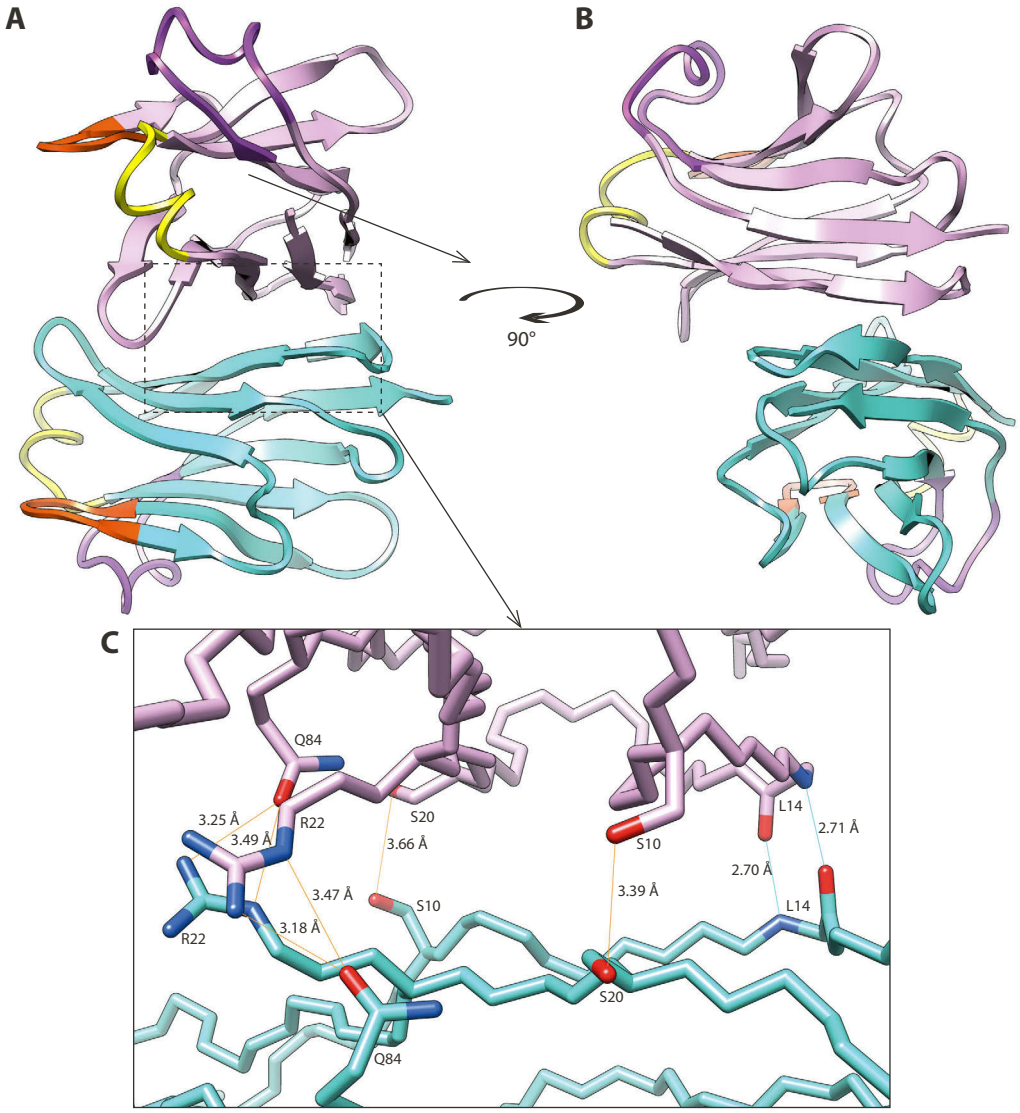


Figure 48 – Structure of a nanobody dimer from the 33SSP–NB33_1 structure. **(A)** The two nanobodies are shown as ribbon models. One nanobody is coloured pink and the other nanobody aquamarine. CDR1 is coloured yellow, CDR2 red and CDR3 purple in both nanobodies. **(B)** 90° rotation of the nanobody dimer. **(C)** Closer look at the dimerisation interface indicated in panel A. The nanobodies are shown as backbone traces. Residues and backbone atoms involved in hydrogen bonding are shown and labelled. Possible hydrogen bonds are indicated with thin lines and the distances between donor and acceptor atoms are indicated.

3. Discussion

Since the start of the research project described in this thesis, two structures of other bestrophin homologues have been published by other groups. From these structures, much has been learned about the function of bestrophins. With the structure presented in this thesis, this knowledge can be further expanded. However, important questions remain that cannot be clearly answered based on these structures alone. In this chapter I will attempt to discuss some of these questions and provide an outlook for further studies.

3.1. Homologue screening and crystallisation

The homologue screening strategy used in this project was successful. Several homologues were identified that could not only be purified, but also crystallised. However, as the homologue screening was done with many homologues in parallel, the focus was solely on identifying well behaving homologues. Therefore, homologues that did not behave well were not tested further. This was also the case for the homologue 67KPN from *Klebsiella pneumoniae*. During the initial screening this homologue behaved promising, but during the later large-scale purification the protein precipitated when the GFP-tag was cleaved off. Consequently, no more attempts were made to purify this homologue since the focus was instead shifted to other homologues. However, some time later a crystal structure of the same homologue (referred to as kpBEST) was published by another group [67]. It is not known why the purification of 67KPN failed, but undoubtedly it would have been possible to purify this homologue if more time had been invested in optimising the expression and purification.

Fortunately, several other homologues behaved well, but they did not all provide well-diffracting crystals. In case of 33SSP, it was generally difficult to reproduce well diffracting crystals. When crystallised either on its own or in complex with the selected nanobody, a large fraction of crystals diffracted poorly. This made it difficult to perform certain experiments, such as the location of potential ion binding sites using anomalous scattering. Although I have attempted to crystallise 33SSP in the presence of ions with anomalous scattering properties such as caesium and bromide, none of the tested crystals diffracted sufficiently well and thus were not used for data collection. Soaking such ions into the crystals did not yield any results either. It is possible that this problem would have eventually been overcome by testing a sufficient number of crystals. Due to limited time, however, other parts of the project were prioritised.

It is still not known why it was difficult to reproduce well diffracting 33SSP crystals. One explanation might be problems during cryoprotection. Since the crystals were grown at very low

concentrations of PEG, a substantial amount of cryoprotectant had to be added to avoid the formation of crystalline ice during the freezing.

Another explanation could be that it was difficult for 33SSP to form crystal contacts due to protein flexibility. Based on analysis of the crystallographic B-factors, the transmembrane part of the structure towards the extracellular side appears to be more flexible than other parts of the structure. In the crystal structure of 33SSP–NB33_1, no crystal contacts involving the extracellular side of the protein could be observed (Figure 49). This is unlike the crystal structures of cBEST1 and kpBEST, where the extracellular regions are directly involved in crystal contacts (not shown). As previously mentioned, the interpretation of the density of 33SSP–NB33_1 is not yet completed, since a certain loop on the extracellular side is not fully built for all 20 monomers. However, based on a preliminary analysis of the weak electron density in this region, in does not seem likely that this loop would extend far enough from the current model to provide any crystal contacts.

If there are indeed no crystal contacts between the extracellular regions, then another explanation could be that the crystal lattice is stabilised by polar interactions between neighbouring detergent micelles [78]. If that is the case, the quality of the crystals could be expected to depend on the size of the detergent micelles, since smaller micelles might not be in contact, while too large micelles would not have sufficient space in the lattice [78]. This could possibly explain why 33SSP did not crystallise in DM, despite appearing to be stable in that detergent. Since lipid was added to the purified protein, it is also possible that the exact amount of lipid that became incorporated into the micelles could vary between different protein purifications.

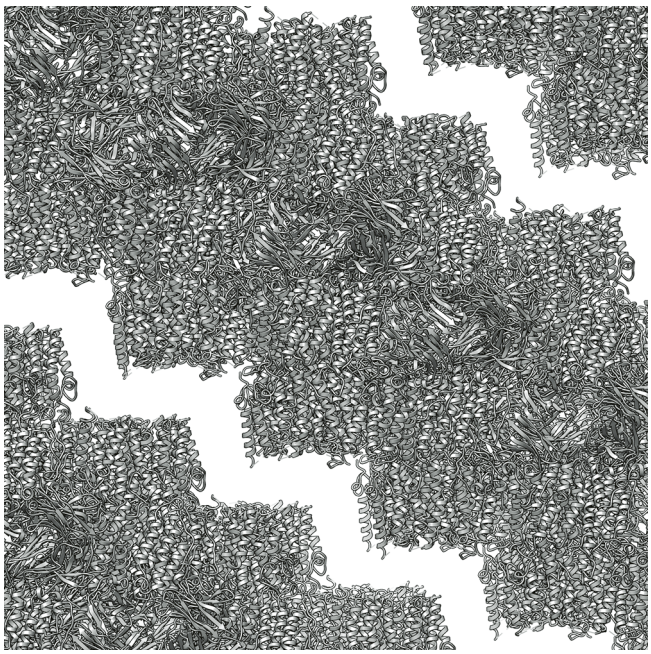


Figure 49 – Crystal packing of 33SSP–NB_1. Several unit cells are shown as grey ribbons. Gaps are observed between extracellular regions of neighbouring molecules.

That would probably also affect the micelle size to some extent. Therefore, part of the reason why the diffraction quality of the tested crystals was variable, could be due to variation in micelle size. If further attempts to crystallise 33SSP are made, it could be investigated if the addition of small amounts of slightly shorter or longer chain detergents, could improve the diffraction quality or reproducibility of the crystals.

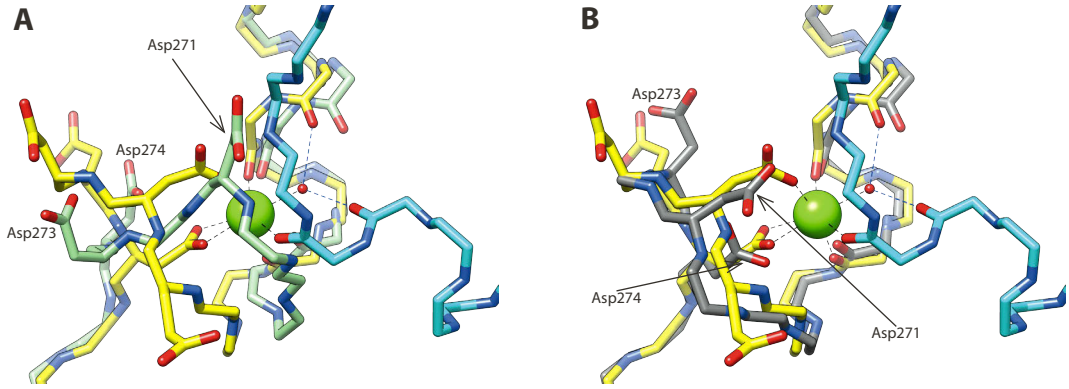
3.2. Structure

3.2.1. Ligand binding

While the motif of acidic residues involved in calcium binding appears to be highly conserved in animal bestrophins, this does not appear to be the case for bacterial homologues. In kpBEST only one of the five acidic residues is conserved, whereas in 33SSP three are conserved (Figure 50c). kpBEST has been reported to be active in the absence of calcium, consistent with the binding site not being conserved [67]. However, in the case of kpBEST the recorded currents were single-channel currents measured in a planar lipid bilayer. No macroscopic currents were recorded and it is debatable if the channel really is fully active in the absence of any ligands or if the recorded currents rather represent low basal activity that occurs in the absence of an unknown ligand.

It thus seems likely that both 33SSP and kpBEST could be activated by some ligand, although the identity of such a ligand is so-far unknown. The two channels could also very well be activated by different ligands since they evolutionally appear to be quite distantly related, with a shared sequence identity of approximately 25%. While this stretch of acidic residues found in animal bestrophins is not conserved in prokaryotic homologues, another feature is conserved. Directly preceding the acidic residues is a highly conserved Pro-Phe-Gly motif. Of the 128 bacterial homologues that were selected for the initial screening, the motif is conserved in 84 homologues (66%), while the remaining 44 homologues (34%) contain Pro-Phe-X. While it is highly speculative, it is possible that bacterial bestrophins are also activated by binding of a ligand in the loop following this conserved Pro-Phe-Gly motif. For that reason, this loop will be referred to as the proposed ligand binding site.

When comparing the individual bestrophin subunits in the 33SSP structure, the proposed ligand binding site is found to exist in two different conformations (Figure 44). Figure 50a–b compares these two conformations to the conformation the calcium binding loop found in cBEST1. The first conformation appears to be collapsed compared to the cBEST1 structure, while the other conformation apparently is in an open conformation which is overall similar to that of cBEST1. Interestingly, the positions of the Asp271 and Asp274 sidechains are close to the positions of Asp301 and Asp304 in cBEST1, which are the two acidic residues that are directly interacting with the calcium-ion in the cBEST1 structure. Three main-chain carbonyls



C

cBEST1	297	P F G E D D D D F E T	307
kpBEST	262	P F G T A A N D L P L	272
33SSP	267	P F G I D S D D L P L	277

Figure 50 – The proposed ligand binding site of 33SSP compared to the Ca^{2+} binding site of cBEST1. The Ca^{2+} binding site of cBEST1 is shown as a yellow and cyan backbone trace, where yellow indicates the region containing the conserved acidic residues, while cyan indicates part of the N-terminus. The bound Ca^{2+} ion is shown as a large green sphere, while a water molecule is shown as a small red sphere. The side chains are shown for all the five highly conserved acidic residues and the main-chain carbonyls that are involved in Ca^{2+} binding are shown as well. Comparisons to 33SSP were made by locally superimposing the shown regions of 33SSP on the structure of cBEST1. For 33SSP, sidechains are only shown for the conserved acidic residues. **(A)** Comparison between the collapsed conformation of the proposed ligand binding site in 33SSP (light green) and cBEST1. **(B)** Comparison between the open conformation of the proposed ligand binding site in 33SSP (dark grey) and cBEST1. **(C)** Multiple sequence alignment of the Ca^{2+} binding sequence of cBEST1 and the same regions from kpBEST and 33SSP. The five, highly conserved acidic residues are highlighted on the cBEST1 sequence.

are in similar positions in the 33SSP structure, compared to those involved in calcium binding in cBEST1. In the cBEST1 structure two N-terminal main-chain carbonyls additionally appear to contribute to the calcium binding, but in the 33SSP structure, the N-terminus is located further away.

As previously mentioned, the five acidic residues found in animal bestrophins appear to be poorly conserved in bacterial bestrophins. In the case of 33SSP, the two acidic residues directly involved in binding calcium in cBEST1 appear to be conserved. It could thus be possible that 33SSP binds calcium or a similar ion, although probably with lower affinity than the calcium binding in animal bestrophins. While kpBEST only has the last of these two acidic residues conserved, 77 out of 128 prokaryotic homologues (60%) selected for screening were found to have the sequence X-Asp/Glu-X-X-Asp. 35 of these homologues were found to have three or four acidic residues located in this loop. As discussed previously, 33SSP and other of the tested homologues, show a higher thermostability in the presence of calcium. This observation does not necessarily mean that calcium binds in this loop, and even if calcium would bind there that would not necessarily lead to activation of the channel. It does, however, remain an interesting observation that should be investigated further.

Another interesting observation comes from a fungal bestrophin homologue from *Aspergillus nidulans*, which has been reported to be calcium dependant, despite only having two acidic

residues in the proposed ligand binding region [66]. Unfortunately, no further studies have been performed with this fungal homologue, but if it is truly calcium activated that could also be the case for 33SSP and other prokaryotic homologues.

3.2.2. Gating

As described previously, the observed conformations of 33SSP, kpBEST and cBEST1 all have a similar minimum pore radius, although this minimum is located at different points along the pore axis. In the case of cBEST1, it has been suggested that the solved structure represents an open or nearly open state [37], since the channel has been crystallised in the presence of calcium, which is known to activate it. As cBEST1 is anion selective, a typical permeant ion would under physiological conditions be chloride. Given an effective ionic radius of chloride of 1.81 Å [125], it is questionable whether the cBEST1 structure really represents an open state.

It has recently been suggested that the residues in the neck of cBEST1 form a gate that can open and close in response to calcium, since mutating the three pore lining hydrophobic residues in the neck to alanines, leads to a channel that is active in the absence of calcium [38]. Hydrophobic gates have also been observed in several other families of ion channels [126]. While the exact sequence of residues in the neck region varies between the available bestrophin structures, they are all hydrophobic and located at similar positions in the different structures. This suggest that at least part of the gating mechanism might be conserved across evolutionary distant homologues.

Despite having several crystal structures available, it is still unknown what kind of structural rearrangements occur during opening of the gate. For animal bestrophins, the opening is assumed to happen upon calcium binding. It is, however, not known how the binding causes opening of the channel and it is also not known what the open channel looks like.

Since the structure of cBEST1 has not been solved without calcium, the conformation of the calcium binding site in absence of calcium is not known either. It is also not known where the cBEST1 N-terminus is located when calcium is not bound. In the cBEST1 structure the calcium ion appears to be buried in the protein, but it would become solvent exposed if the N-terminus would move away [37]. If the N-terminus would indeed change position upon the binding of calcium, this could also mean that the first transmembrane helix would change position, since this is located directly following the short N-terminal region. In this way binding of calcium could lead to a structural change, which could cause opening of a gate.

Another structural change that could possibly happen upon ligand binding can be observed in the structure of 33SSP. As mentioned, the so-called 140-loop appears to be flexible and can change between different conformations in order to accommodate the open conformation of the proposed ligand binding site. If the gate is placed in the transmembrane region, it would

be hard to explain how a slight movement of this seemingly flexible cytoplasmic surface loop would lead to structural rearrangements in the transmembrane region. It does, however, remain a possibility that should be kept in mind for further studies.

3.2.3. Selectivity

While it seems likely that the gating of ion flow is performed in the hydrophobic neck region, it is so far unknown how selectivity for different ions arises. In the case of cBEST1, the aperture has been suggested to form a size-selective filter, allowing easier passage to larger anions that are more easily dehydrated [38]. This is based on the observation that mutating the valine residue, which constricts the aperture of cBEST1, to alanine results in anions of different sizes permeating equally well. Since this mutation or the previously mentioned triple-alanine mutation in the neck region do not change the anion over cation selectivity of cBEST1 [38], it is so far unknown what causes cBEST1 to be an anion channel and kpBEST a cation channel [67].

In cBEST1 it has been proposed that the anion selectivity could be due to anion binding sites located just outside the neck and in the inner cavity of cBEST1, where chloride ions were observed in the structure [37]. These binding sites appear similar to those found in the glutamate-gated chloride channel α (GluCl) from *Caenorhabditis elegans* [127], although the binding sites in that protein are on the intracellular side of the pore. Interestingly, GluCl belongs to the family of Cys-loop receptors, a group of pentameric ligand gated ion channels, which also includes cation selective channels such as nicotinic acetylcholine receptors. In cation selective Cys-loop receptors, a ring of negative charges is typically found at the cytoplasmic side of the pore [128,129]. This ring of negative charges seems to affect the charge selectivity both in eukaryotic [130] and prokaryotic family members [131], although other mutations are also required to fully convert a cation selective channel to an anion selective one [132].

As described earlier, rings of negative charges were also observed at different positions in the cytoplasmic region of 33SSP. At present the charge selectivity of 33SSP is unknown, but since the other crystallised bacterial bestrophin homologue, kpBEST, appears to be cation selective, this could potentially also be the case for 33SSP.

Although the charge selectivity of 33SSP is not known, it does seem questionable if a negatively charged ion would be able to pass through the negatively charged aperture. Since 33SSP also appears to lack positive binding pockets outside the pore, a plausible hypothesis could be that 33SSP forms a cation selective channel like kpBEST. This is of course still speculative, since the mechanism for ion selectivity in cBEST1 and kpBEST is not fully understood.

When comparing the 33SSP sequence to those of other bacterial homologues, the Asp82 position was found to be conserved. While it is not present in kpBEST, out of the 128 homologues that were selected for screening, 100 of them (78%) had either an aspartate or glutamate resi-

due at this position. By comparison, in a set of 487 chordate bestrophin sequences downloaded from the UniProt database [133], none of the aligning sequences were found to have an acidic residues at this position.

The picture is, however, complicated by the previously discussed fungal homologue from *Aspergillus nidulans*. This homologue has been reported to form an anion selective channel [66]. When aligning the sequence of the fungal homologue to the sequence of 33SSP, it appears that the fungal homologue has a glutamate residue at the position of Asp82. However, since this fungal homologue appears to be only remotely related to bestrophins from both animals and prokaryotes, it is difficult to judge whether this residue would have a similar conformation in the structure. Since the relationship of the fungal homologue appears to be distant, most of its sequence cannot be aligned well to the sequences of cBEST1, kpBEST or 33SSP. The sequence around Asp82 seems to align reasonably well, but no conclusions can be made about the other two acidic residues found along the pore of 33SSP.

In prokaryotes, these other two acidic residues appear to be much less conserved than Asp82. 40 out of 128 (31%) prokaryotic homologues have an acidic residue at the position of Glu185, while 28 out of 128 (22%) have an acidic residue at the position of Asp192. At the aperture, none of the examined chordate sequences appear to have an acidic residue at the position of the 33SSP Glu185. However, at the position equivalent to Asp192, some of the chordate sequences, including human BEST2 and human BEST4, do contain an acidic residue, so it is probably less likely that this residue is of critical importance for the charge selectivity, compared to the two other discussed residues.

Whereas a high degree of sequence conservation does not necessarily define its functional importance, there is still an important correlation. Given that an acidic residue at the position of Asp82 appears to be highly conserved in prokaryotic bestrophin sequences, but completely absent from chordate bestrophin sequences, this could suggest a functional role of this residue in bestrophin selectivity, given that the selectivity of prokaryotic homologues differs from animal bestrophins. So far, the assumption that prokaryotic bestrophins might be cation selective is only based on one study of kpBEST. That study was, however, not very thorough, so more work is required to better understand the selectivity of prokaryotic bestrophins.

3.3. Summary and outlook

Several questions regarding the function of 33SSP and other prokaryotic bestrophins are still not answered, as it has been discussed above. It is not yet known how prokaryotic bestrophins, including 33SSP, are activated. Also, it is not known how the binding of a ligand can cause the opening of channel. Finally, the selectivity of prokaryotic bestrophins is not fully understood either.

As discussed in the appendix, some initial experiments have been performed with reconstituted 33SSP. Attempts were made to fuse liposomes containing the reconstituted protein to planar lipid bilayers. Although currents could be observed, it is at this stage, however, not conclusive that those currents were produced by 33SSP.

Further studies should therefore focus on understanding the function of 33SSP using other methods. The crystal structure presented in this work may serve as a valuable framework for such studies, since the choice of experiments can now be based on structural knowledge. Hopefully this will lead to a better understanding of the structure and function of the bestrophin protein family.

4. Materials and Methods

4.1. Bioinformatics

Sequences of various prokaryotic bestrophin homologues were downloaded from the NCBI database [68]. For the second round of screening, sequences were aligned in the program CLC Main Workbench [134], using the default algorithm, followed by construction of a Neighbour Joining (NJ) tree, also using CLC Main Workbench. For the third round of screening, homologues were aligned using Clustal Omega [135]. Sequences that aligned poorly were manually removed. Columns in the alignment that mainly contained gaps were also removed. Based on the alignment, a maximum likelihood tree was constructed using PhyML [136].

For analysis, smaller sets of sequences were aligned using PROMALS3D [137] whenever structural information was available. The full set of 128 bacterial homologues selected for screening, were aligned using MAFFT [138] (using the E-INS-i method, with default settings), before constructing a maximum likelihood tree with PhyML. The MAFFT alignment was furthermore used to quantify the number of conserved residues in the proposed ligand binding site. A set of 487 chordate bestrophin sequences was downloaded from UniProt [133] and aligned with MAFFT (using the FFT-NS-i method, with default settings). Alignments were visualised using Jalview [139].

4.2. Cloning of homologues

For the first round of screening, all prokaryotic bestrophin homologues were cloned from genomic DNA prepared for another project [77]. For the second and third round of screening, most homologues were cloned from genomic DNA that was either purchased directly from the German Collection of Microorganisms and Cell Cultures (DSMZ) or purified from bacterial pellets, which were also purchased from DSMZ. Purification of genomic DNA was performed with the GenElute Bacterial Genomic DNA kit (Sigma-Aldrich). When genomic DNA or bacterial pellets were not available at DSMZ (homologues 93–94 and 108–128), genes were synthesised in a codon optimised version by GenScript.

All cloning was done using the FX-cloning system [140], using the recommended primer design and cloning procedures. PCR was performed essentially under standard conditions using Phusion polymerase (New England BioLabs). The main exception from the standard conditions was that the polymerase was added when the reaction mixture had reached a temperature of 90°C, before the initial denaturation step. Homologues were initially cloned into

the FX-cloning compatible entry vector pINITIAL_cat (chloramphenicol resistant) and then sequenced. After sequence verification, the homologues were sub-cloned to an FX-cloning compatible, L-arabinose inducible expression vector, such as pBXC3GH (ampicillin resistant), which added a HRV 3C cleavage site, GFP and 10×His-tag to the C-terminus of the inserted protein. All cloning and expression was performed with the *E. coli* strain MC1061. Terminal truncations of different homologues were made with FX-cloning compatible primers that only amplified a limited part of the gene.

4.3. Small scale expression screening

One day before expression, overnight cultures were prepared in Terrific Broth medium (TB) containing ampicillin. On the next day, the overnight cultures were used to inoculate fresh TB (including ampicillin). Typically, cultures were either grown as 1 ml or 4 ml cultures in deep well plates (96 or 24 wells, respectively). Cultures were incubated in an orbital shaker for 1.5 hours at a temperature of 37°C. The temperature was then reduced to 25°C over the course of 1 hour. Expression was induced by addition of L-arabinose to a final concentration of either 0.04%, 0.004% or 0.0004%. In the later rounds of screening, the concentration 0.004% was used exclusively. After addition of L-arabinose, growth was continued overnight at 25°C. In the initial screening round, the lactose transporter LacS (in a truncated form lacking the last 160 amino acid residues [141]) from *Streptococcus thermophilus* was included as a positive control. In later screening rounds the prokaryotic bestrophin homologues 12CPI or 33SSP were used instead of LacS.

On the following day, cells were harvested by centrifugation. All following steps were done with cold buffers and either on ice or in a cold room (4°C) to the extent that it was possible. Cells were resuspended in PBS. The optical density at 600 nm (OD_{600}) as well as the whole-cell GFP fluorescence (excitation at 485 nm; emission at 535 nm) was measured using an Infinite M1000 (Tecan) plate reader. Fluorescence readings were normalised to the number of cells as judged by the OD_{600} reading.

In 2 ml screw cap tubes, cells were added in an amount equivalent to 175 µl with an OD_{600} value of 1. After a brief centrifugation, the cells were resuspended in 400 µl of lysis buffer (50 mM potassium phosphate (KP_i), pH 7.5, 1 mM $MgSO_4$, 10% glycerol, 1mM PMSF, and 3 µg/ml DNase I; for the last round of screening additionally 150 mM NaCl). Approximately 300 mg of glass beads (diameter of ≤106 µm; Sigma) were added to each tube. Cells were lysed by shaking intensely in a FastPrep-24 benchtop homogeniser (MP Biomedicals) for 2 × 20 s at a speed of 6 m/s. Glass beads were removed by gentle centrifugation and DDM was added to a concentration of 1%. After 1 hour incubation on ice, 120 µl of the detergent extract was cleared

by ultracentrifugation for 15 min at 75,000 RPM, using the rotor TLA-100 (Beckman Coulter) placed in the Optima MAX-XP ultracentrifuge (Beckman Coulter).

Samples from before and after the ultracentrifugation step were analysed by SDS-PAGE. 8 µl sample was mixed with 2 µl 5× Sample Buffer [74] and loaded on a gel. After completion of the electrophoresis, the GFP fluorescence was visualised directly in the gel by using a LAS-3000 Imaging System (Fuji), which illuminated the gel with 460 nm blue light and filtered the emitted light through a Y515 filter.

After recording the GFP fluorescence, the gel was analysed by western blot. The proteins were transferred to an Immobilon-P PVDF membrane (Merck Millipore), using the semi-dry blotting procedure. The membranes were blocked with 5% skimmed milk powder in PBS and then incubated with an anti-His_6-peroxidase antibody (Roche). ECL substrate (Biorad) was added and signals detected with the LAS-3000 Imaging System (Fuji).

In parallel with the SDS-PAGE analysis, 50 µl of the sample from after the ultracentrifugation was analysed by fluorescence-detection size-exclusion chromatography (FSEC). The sample was injected on a Zorbax GF-450 column (Agilent) connected to an Agilent 1100 HPLC system with fluorescence detection (excitation at 480 nm, emission at 530 nm). The column was equilibrated in 50 mM potassium phosphate (KP_i), pH 7.5, 150 mM NaCl, 0.03% DDM for the first two rounds of screening. For the last round, the buffer was 20 mM HEPES pH 7.5, 150 mM NaCl, 0.03% DDM. Terminal truncations of different homologues were evaluated using the same FSEC procedure.

4.4. Large scale expression and preparation of membrane vesicles

One day before expression, overnight cultures were prepared in TB containing ampicillin. On the next day, the overnight cultures were used to inoculate fresh TB (including ampicillin). Cells were either grown as multiple (typically 32) 0.6 L cultures in 2 L shaker flasks, or as 9–18 L cultures in a fermenter (Bioengineering). The initial temperature of 37°C was after approximately 1.5 hours lowered to 25°C. Once the OD_{600} had reached 1 (shaker cultures) or 3 (fermenter cultures), expression was induced by addition of L-arabinose to a final concentration of 0.004–0.01%. For shaker cultures, growth was continued overnight at 25°C, whereas the temperature of fermenter cultures was lowered to 20°C before being left to continue growth overnight.

On the following day, cells were harvested by centrifugation and resuspended in cold PBS at an OD_{600} of roughly 200. 0.1 mg/ml lysozyme, 2 µg/ml DNase I and 1 mM $MgSO_4$ were added. After one hour of incubation at 4°C, the cells were lysed using a pressure based cell lyser

(custom build by the in-house workshop). Directly after lysis, 1 mM PMSF, 1.5 µM Pepstatin A (Applichem) and 2 µM Leupeptin (Applichem) were added. The lysate was centrifuged for 1 hour at 42,000 RPM in a 45 Ti ultracentrifuge rotor (Beckman Coulter) cooled to 4°C. The supernatant was discarded. The pellet, containing the crude membrane vesicles, was resuspended in PBS + 10% glycerol using a Potter-Elvehjem homogeniser. Finally, the resuspended membrane vesicles were flash frozen in liquid nitrogen, before storage at −80°C.

4.5. Membrane protein purification

All purification steps were done with cold buffers and either on ice or in a cold room (4°C). DDM was mostly used as detergent for the described experiments. In case DDM was replaced by either UDM or DM, the buffers contained 0.09% UDM or 0.2% DM instead of 0.03% DDM. For the detergent extraction and initial steps of the purifications, detergent of the purity Solgrade (<5% α anomer) from Anatrace was used. SEC and later steps were done with Anagrade detergent (<2% α anomer), also from Anatrace.

Membrane vesicles were thawed and diluted to 0.1 g/ml in PBS + 10% glycerol. DDM was added to a concentration of 1.25%. The solution was stirred for 1 hour. Insoluble material was removed by ultracentrifugation. Ni-NTA resin was added to the supernatant, imidazole (pH 7.5) was added to 15 mM concentration and the 10×His-tagged GFP fusion protein was allowed to bind to the resin while mixing for 1 hour. Ni-NTA resin was collected in a column and washed with at least 25 column volumes of wash buffer (20 mM HEPES, 50 mM imidazole, pH 7.5, 150 mM NaCl, 0.03% DDM) and eluted with elution buffer (20 mM HEPES, 400 mM imidazole, pH 7.5, 150 mM NaCl, 0.03% DDM). The eluted protein was mixed with HRV 3C protease in a molar ratio of 1:5 (protease/protein). While cleaving, the sample was dialysed for two hours against a dialysis buffer consisting of 10 mM HEPES, pH 7.5, 150 mM NaCl and 0.03% DDM. After the dialysis, the 3C protease (His-tagged) and the cleaved-off GFP (His-tagged) were removed by binding to Ni-NTA resin. The flow-through from the column, containing the protein of interest, was concentrated using an Amicon Ultra centrifugal concentrator (Merck Millipore) with a 100 kDa molecular weight cut-off. The concentrated sample was injected on a Superdex 200 column (GE Healthcare), typically equilibrated in 10 mM HEPES, pH 7.5, 150 mM NaCl and 0.03% DDM. If the protein was intended to be used for crystallisation experiments, 5 mM HEPES was used instead of 10 mM. The peak fraction was analysed by SDS-PAGE. The absorbance at 280 nm was measured and concentrations were calculated with extinction coefficients calculated by the ExPASy server [142]. Depending on the application, the purified protein was further concentrated using a centrifugal concentrator.

4.6. Lipid addition

Lipids were initially added to the concentrated protein from a stock of liposomes in buffer. Extra detergent was added as needed from a separate stock. Later the HiLiDe [82] protocol was adopted as follows. Lipid solutions in chloroform was pipetted into small glass tubes (75 mm long, 8 mm inner diameter). For most experiments *E. coli* polar lipids or DOPC was used (both from Avanti Polar Lipids). Chloroform was evaporated under a nitrogen stream and the tubes were placed in a vacuum desiccator for at least 2 h to remove residual chloroform. Concentrated protein was added directly to the glass tube along with detergent from a 10% stock solution. The final lipid concentration was in the range 0.1–0.25%, but typically 0.15% was used. Detergent concentration was in the range 0.5–1.0%, with 0.75% being most typically used. Typically, the final volume of the sample was at least 100 µl. A small PTFE-coated stirrer bar (5 × 2 mm) was added and the mixture was stirred overnight (130 RPM) at 4°C. On the next morning, non-solubilised lipids were removed by ultracentrifugation (10 minutes at 190,000 g), before the sample was used for crystallisation.

4.7. Crystallisation, data collection and data processing

Initial crystallisation screening was carried out by an in-house core facility. Protein drops (typically 0.1 or 0.2 µl) were mixed with an equal volume of reservoir solution in a 96-well sitting drop plate. The screen used was a custom grid screen consisting of 6 × 96 different conditions. Various salts were combined with different buffers at 6 different pH values ranging from 4.5 to 9.4. For each buffer/salt combination, 4 different concentrations of PEG400 (15%, 25%, 35% and 45%) and 4 different PEG4000 concentrations (5%, 10%, 20% and 30%) were tested. A later version of the screen used exclusively PEG400, which was tested at 8 different concentrations (starting from 10%, increased in 5% steps). Crystallisation plates were incubated at 4°C or 20°C and automatically imaged by a Crystal Farm (Bruker) or a Rock Imager 1000 (Formulatrix). Conditions where crystals were observed were typically refined with 1 µl drops that were set up by manual pipetting in 24 well Cryschem M plates (Hampton Research). Typically, PEG concentrations increasing in 1% increments were used.

Crystals were cryoprotected by increasing the PEG400 concentration in 5% steps, until a concentration of 35% was reached. For crystals growing in PEG4000, ethylene glycol was used as a cryoprotectant. Crystals were frozen in liquid propane. Most data was collected at the Swiss Light Source (SLS), at beamline X06SA on a Pilatus 6M detector. For the initial screening, occasionally a Mar225 CCD detector was used.

Data was indexed, integrated, scaled and merged using the XDS package [98]. The space group was determined using the program POINTLESS [99], part of the CCP4 suite [143]. Data statistics were calculated using the phenix.merging_statistics, part of the PHENIX suite [109].

4.8. Multi-angle light scattering

Protein for multi-angle light scattering experiments was purified as described previously, except that only the central part of the size-exclusion peak was used for analysis. 50 µg of purified protein at a concentration of 1 mg/ml was injected on a Superdex 200 column, equilibrated in the same buffer as used for purification by size-exclusion chromatography. The column was connected to an Agilent 1100 HPLC system with a UV detector. After the UV detection, the sample was further analysed using miniDAWN TREOS MALS detector and an Optilab T-rEX refractometer (both from Wyatt Technology). Bovine serum albumin was used as a protein standard and data was analysed using the Astra software suite (Wyatt Technology). Measurements were done at 20°C.

4.9. Alpaca immunisation and generation of nanobody library

33SSP was purified as described previously. 4 injections of 200 µg protein were injected into to an alpaca over the course of 6 weeks. The animals were not harmed and less than 100 ml of blood was collected in the final blood draw. All alpaca related work was done by Dr. Saša Štefanić (Vetsuisse Faculty, University of Zurich), who had acquired permission for the experiment from the Cantonal Veterinary Office (Zurich, Switzerland).

The following steps were performed by Yvonne Neldner. Leucosep tubes (Greiner Bio-One) were filled with Ficoll Paque Plus (GE Healthcare). Anticoagulated alpaca blood was diluted 1:1 with 154 mM NaCl. 15–30 ml diluted blood was added to each Leucosep tube, which was centrifuged for 10 minutes at 1000g in a swing-out rotor at room temperature. Peripheral mononuclear cells were removed with a Pasteur pipette and washed twice with PBS. Total RNA was isolated using the RNeasy Mini Kit (Qiagen). First-strand cDNA was synthesised using the AffinityScript qPCR cDNA synthesis kit (Agilent) in combination with the CALL002 primer [144]. VHH (nanobody) sequences were amplified by PCR using the forward primer CTG ARC TKG GTG GTC CTG GCT GC and a 1:1 mix of the reverse primers CTG ARC TKG GTG GTC CTG GCT GC and TTC GGN GGG AAG AYR AAG AC. A fragment of approximately 600 bp was isolated by gel electrophoresis and purified from the gel. The isolated fragment was further amplified by PCR using the FX primers ATA TAT GCT CTT CGG CAC AGN KGC ARY TSG TRG AGT CTG GGG G and ATA TAT GCT CTT CCA CTG GAG ACG GTG ACC SGG GTC CCY TKG CYC. A fragment of approximately 400 bp was isolated by gel electrophoresis and purified from the gel. An FX cloning compatible version of the vector pMES4 [145] was linearised using the restriction enzyme PmlI and dephosphorylated with calf intestinal alkaline phosphatase. The isolated DNA fragment was cloned into the linearised vector, using the FX cloning procedure, to generate the VHH library. Electro-competent *E. coli*

TG1 cells were transformed with the VHH library by electroporation. A small amount of the cells was used for a dilution series, in order to calculate the library size. The remaining cells were plated and grown overnight. The next day the cells were harvested from the plates and used to prepare glycerol stocks.

4.10. Selection of nanobodies using phage display

33SSP was expressed with a C-terminal Avi-tag [92] fused to the C-terminus, followed by a 3C cleavage site, GFP and 10×His-tag. 33SSP was biotinylated in a reaction mixture consisting of 30 mM 33SSP, 3 mM biotin, 5 mM ATP, 10 mM MgOAc, 20 mM HEPES pH 7.4, 0.1 mg/ml BirA (*E. coli* biotin ligase) and DDM. The reaction was incubated overnight at 4°C. On the next day, biotinylated 33SSP was purified using size-exclusion chromatography as previously described.

Two selections were performed in parallel. In one selection neutravidin was used to capture biotinylated 33SSP (neutravidin coating) and in another selection 33SSP was directly coated on the plastic (direct coating). 100 µl neutravidin (4 µg/ml) in 0.1M $NaHCO_3$ pH 8.2 or 33SSP (2 µg/ml) in PBS + DDM was coated in a MaxiSorp 96-well ELISA plate (Nunc). The wells were blocked with 200 µl 2% milk powder in PBS + DDM. To the plate coated with neutravidin, 100 µl biotinylated 33SSP (2 µg/ml) in PBS + DDM + 0.2% milk powder was added after washing. For the first round, one negative control well was prepared by blocking with 2% milk powder in PBS. For the subsequent rounds, two negative controls were prepared. One contained only milk powder (control for the direct coating selection). The other was first coated with neutravidin and then blocked with 2% milk powder (as a control for the neutravidin coating selection).

For the first round of selection, 100 µl of phages at a concentration of 1.4×10^{11} phages/ml (in PBS + DDM + 1% milk) was added to each well. After 2 hours of incubation, each well was washed 15 times with PBS + DDM. The phages were eluted by incubating with 100 µl 0.25 mg/ml trypsin for 30 min at room temperature. The reaction was stopped by addition of 4-benzenesulfonyl fluoride hydrochloride (0.25 mg/ml). 50 µl of the eluted phages were used to infect 350 µl *E. coli* TG1 cells (freshly cultured from a minimal media plate, in the exponential phase). Infection was done for 30 min at 37°C without shaking. The infected cells were then grown overnight in 2×YT + ampicillin + 2% glucose. A small amount of the eluted phages was used to make a dilution series. 10 µl of diluted phages were similarly used to infect 90 µl *E. coli* TG1 cells, which were plated on LB agar plates (with ampicillin and 2% glucose) for overnight growing.

On the next day, the phage output was calculated from the number of colonies observed from the dilution series. The liquid culture was used to prepare a glycerol stock and to start a fresh

culture in 2×YT + ampicillin + 2% glucose. The culture was grown to the exponential phase and then superinfected with M13KO7 helper phage (prepared in advance, using essentially the same method as described here) at a multiplicity of infection of 20. The phages were allowed to infect for 30 min at 37°C without shaking. The cells were then pelleted and resuspended in 50 ml 2×YT + ampicillin + kanamycin and grown overnight at 37°C with shaking. On the following day, the cells were removed by centrifugation and 40 ml of the supernatant was mixed with 10 ml 20% PEG6000, 2.5M NaCl. The sample was mixed by inverting the tube and stored on ice for 30 min. The phages were pelleted by centrifugation, resuspended in 1 ml PBS, transferred to a 1.5 ml tube and mixed with 250 µl 20% PEG6000, 2.5M NaCl. The sample was again mixed by inversion and stored on ice for 15 min. The phages were collected by centrifugation and resuspended in 1 ml PBS. Aggregated particles were removed by centrifugation and the supernatant, containing the phages, was transferred to a new tube. A small amount of the phages was used for a dilution series, as described above, to determine the phage concentration. A part of the remaining phages was used for the next round of selection or frozen for later use.

4.11. ELISA screening and sequencing of single nanobody clones

For ELISA analysis, glycerol stocks prepared from the output phages were streaked on LB-agar (with ampicillin and 2% glucose) and grown overnight. On the next day, colonies were picked. Each colony was first transferred to a fresh patch plate and next to transferred to 1 ml TB + ampicillin in 96-well deep well plate, where it was grown for 6 h at 37°C. The temperature was then reduced to 25°C and each culture was induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside. Cells were grown overnight at 25°C.

On the next day, cells were pelleted at 4°C and re-suspended in 200 µl cold TES buffer (0.2 M Tris-HCl pH 8.0, 0.5 mM EDTA, 0.5 M sucrose) by vortexing. The cells were shaken for 30 min in the cold room, before 300 µl cold TES/4 buffer (TES buffer diluted 4× in water) was added. After mixing, the cells were shaken again for 30 minutes in the cold room. Cells were pelleted and 400 µl of the supernatant, containing the periplasmic extract, was transferred to a new plate, where it was mixed with 100 µl PBS + 0.25% DDM.

ELISA plates were coated either directly with 33SSP (for analysis of clones selected using the direct coating procedure) or with neutravidin followed by biotinylated 33SSP (for analysis of clones selected using the neutravidin coating procedure), as described previously. Every second column in the ELISA plate was used for negative controls, meaning that no 33SSP was added. After blocking with 2% milk powder in PBS, the periplasmic extract was added. After 1 h incubation at room temperature, the plate was washed with PBS + 0.05% DDM. An

anti-His_6-peroxidase antibody (Roche) was diluted 1:2500 and added. After washing, signals were detected by adding BM Blue POD substrate (Roche). After sufficient colour had developed, the reaction was stopped by addition of 1 M H_2SO_4 and the absorbance at 450 nm was measured using an Infinite M1000 (Tecan) plate reader. For the clones selected for sequencing, a culture was started from the patch plate, plasmid DNA was isolated and the insert was sequenced. For all sequences, vector encoded elements were removed and the sequences were aligned.

4.12. Expression of nanobodies, complex formation and crystallisation

Selected nanobodies were cloned into the FX-cloning compatible, L-arabinose inducible expression vector pBXNPHM3 (ampicillin resistant), which added a PelB leader sequence, MBP, 10×His and a HRV 3C cleavage site to the N-terminus of the nanobody. Expression with the *E. coli* strain MC1061 and cell lysis, were performed as described previously for the bacterial bestrophin homologues. After ultracentrifugation, the supernatant was mixed with Ni-NTA and the purification proceeded as for the bacterial bestrophin homologues, except that a Superdex 75 column was used for the gel filtration step and that no detergent was added in the various buffers. Purified nanobodies were used immediately or stored at -80°C for later use.

For a test of complex formation, purified nanobodies were mixed with purified 33SSP. The final concentrations were 1 mg/ml of nanobody and 2.5 mg/ml of 33SSP, resulting in an approximately 1:1 molar ratio. Complexes were analysed by SEC, using a Zorbax GF-450 column (Agilent) connected to an Agilent 1100 HPLC.

When 33SSP–nanobody complexes were prepared for crystallisation, purified nanobodies were added in a slight molar excess to 33SSP before the SEC-step. Purification then proceeded as usual for 33SSP. Addition of lipids and crystallisation was performed as described previously for the bacterial bestrophin homologues.

For crystallisation of 33SSP–NB33_1, protein was incubated overnight at a concentration of 9 mg/ml together with 0.15% *E. coli* polar lipids and 0.75% DDM. The buffer used was identical to the size-exclusion buffer (150 mM NaCl, 5 mM HEPES, pH 7.5) On the next day, after clearing the sample by ultracentrifugation, EGTA was added to a concentration of 2 mM. Crystallisation plates were set up with sitting drops consisting of 1 µl of protein and 1 µl of reservoir solution. The two solutions were mixed by pipetting up and down. The reservoir solution consisted of 13% PEG300 (v/v), 50 mM ADA, pH 6.5 (pH adjusted with NaOH). Plates were sealed with tape and incubated at 4°C. Crystals appeared within a few days and were typically frozen within a few weeks. The crystals were cryoprotected by transfer into solutions containing increasing concentrations of PEG300 (18%, 24%, 30% and 35%). The cryoprotection solutions

contained the same salts and buffer concentrations as the crystallisation drop, low concentrations of detergent and no added lipids. Crystals were frozen by plunging into liquid propane. The best diffracting dataset of 33SSP–NB33_1 was collected using a Pilatus 6M detector at the beamline X10SA of the Swiss Light Source (SLS). The detector was placed at a distance of 550 mm and a wavelength of 1.00002 Å was used. Data was processed as described earlier. 5% of the reflections were randomly selected and assigned to the test set, which was excluded from the refinement.

4.13. Structure determination and analysis

A molecular replacement search model was generated from structure of the bestrophin homologue kpBEST [67] (PDB accession number: 4WD8). Based on a sequence alignment, the search model was adjusted with the program SCULPTOR [106] using default settings. Molecular replacement was performed with PHASER [105], accessed through the CCP4 software suite [143]. 4 33SSP pentamers were searched for. Solvent content was calculated based on 4 33SSP pentamers and 20 nanobodies. RMSD between the search model and unknown structure was specified as 1 Å. The setting «pack compact» was set to «off». Data up to a resolution of 3.5 Å was used. The molecular replacement solution was improved by iterated local density-guided model deformation and refinement («morphing») using the program phenix.morph_model [110].

Based on a structure of a nanobody [77] (part of PDB accession number: 5M94), a molecular replacement search model for the nanobody NB33_1 was prepared, again based on an alignment using the program SCULPTOR. 20 nanobodies were added into the density with the Spherically Averaged Phased Translation Function and Phased Rotation function of the program MOLREP [112]. The model was improved through iterative cycles of building in the program Coot [119] and refinement in the program phenix.refine [108]. Atomic coordinates and isotropic B-factors were refined using automatically determined torsion-angle NCS restraints. Later rounds of refinement also included TLS parameters (one TLS group per chain).

Most figures were prepared with the program UCSF Chimera [146]. Structures were superimposed using secondary structure matching [118] in Coot. Protein interfaces were analysed with PISA [117]. Pore dimensions were calculated with the program HOLE [120], using the default atomic radii. Electrostatic surface plots were made in PyMol (www.pymol.org) using APBS [147], with charges being assigned by using the AMBER forcefield in PDB2PQR [148].

5. Appendix

5.1. Planar lipid bilayer experiments

5.1.1. Introduction

The presented structure of 33SSP shows what appears to be an ion channel in a closed conformation. As previously discussed, not much is known about the function of prokaryotic bestrophins, since the only available functional data comes from an article describing the structure of kpBEST [67]. It was therefore my goal to investigate the function of 33SSP. Since this protein can be purified and is stable in detergent, I decided to reconstitute the protein into proteoliposomes and attempt to record currents in a planar lipid bilayer (in the following referred to simply as 'bilayer'). Bilayer recordings have been used to study both kpBEST and cBEST1 [38,67]. Unlike typical patch clamp experiments on living cells, endogenous currents from other channels can effectively be eliminated in a bilayer system since purified protein is used. Depending on the amount of proteoliposomes that are fused to the bilayer and the protein density of the proteoliposomes, single-channel currents or macroscopic currents can be measured.

As described below, it was possible to measure currents when adding proteoliposomes containing 33SSP to a bilayer. These currents unfortunately appear to be (at least partially) caused by artefacts and not by incorporated 33SSP.

5.1.2. Results

After adding proteoliposomes to the bilayer, macroscopic currents could be recorded. The current amplitude ranged from a few hundred pA to a few nA. To determine if the currents were cation or anion selective, different salt concentrations were used on the two sides of the bilayer and the currents were measured at different voltage steps (Figure 51).

Based on the recordings, current-voltage curves (I-V curves) could be drawn and a reversal potential of –53.4 mV could be determined in a 10-fold gradient of NaCl (Figure 52). Using the Goldman-Hodgkin-Katz voltage equation, a permeability ratio between chloride and sodium ions (P(Cl)/P(Na)) of 0.022 could be calculated, meaning that the observed currents were highly cation selective.

The measurements were performed with proteoliposomes reconstituted at a lipid to protein ratio (LPR) of 50. Similar macroscopic currents were also observed with proteoliposomes pre-

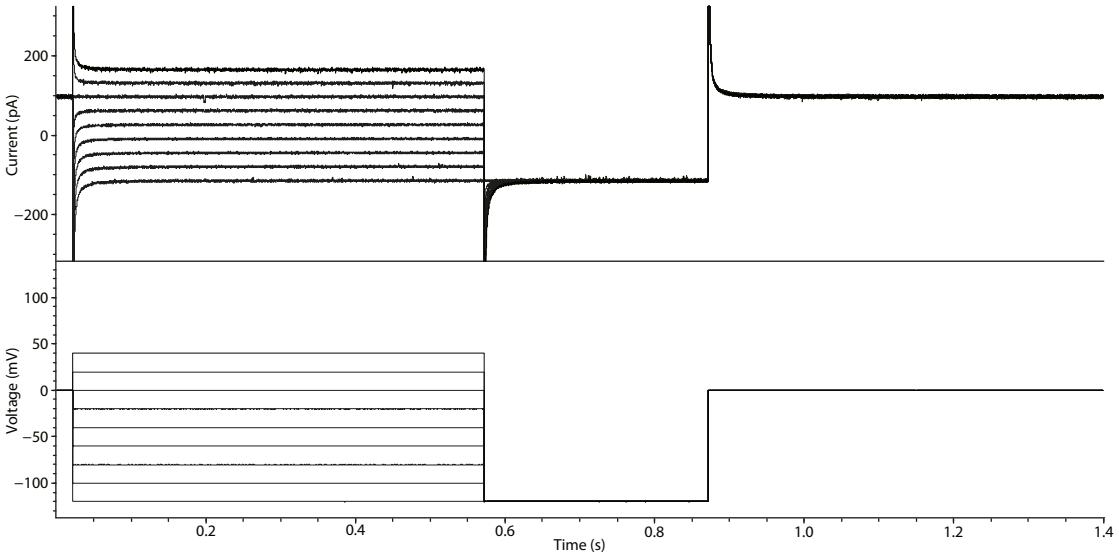


Figure 51 – Example of macroscopic currents recorded at different voltages. The upper part of the figure shows the recorded currents, while the lower part shows the applied voltage steps. The experiment was performed using ten times higher concentration of NaCl on one side compared to the other side of the bilayer.

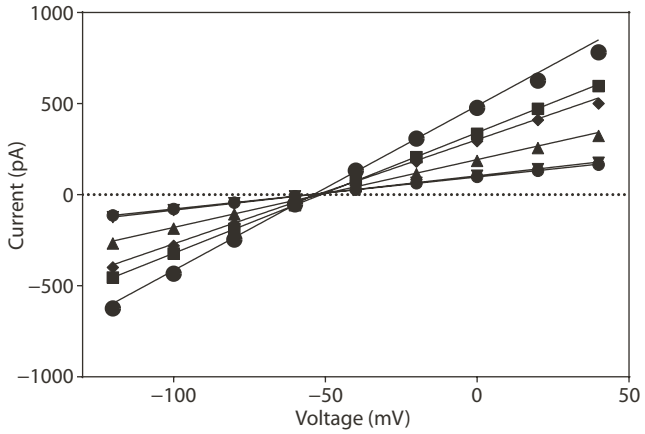


Figure 52 – I-V curves used to determine the reversal potential. The results of six independent experiments are shown, all performed using ten times higher concentration of NaCl on one side compared to the other side of the bilayer. The average reversal potential was determined as –53.4 mV (SD = 1.7 mV).

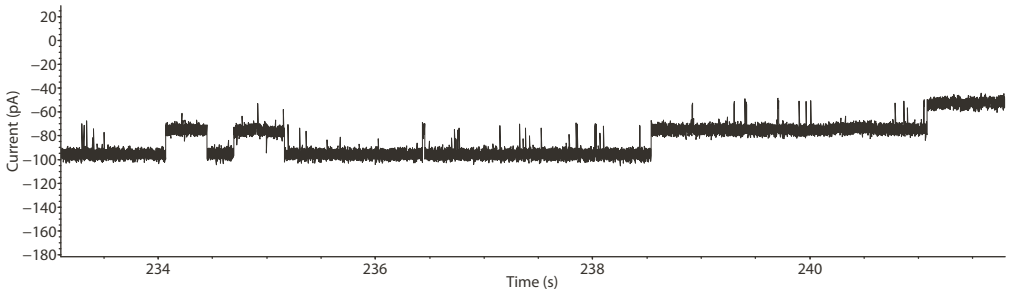


Figure 53 – Example of observed single channel currents. The measurement was performed at –100 mV. The buffers on both sides of the bilayer contained 200 mM NaCl. The observed current transitions correspond to 21 pA or 210 pS.

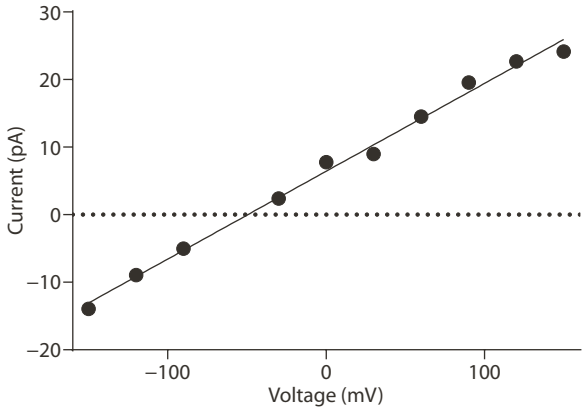


Figure 54 – Cation selectivity of single channel currents. The current amplitudes of the observed single channel transitions were measured at different voltages and an I-V curve was plotted. The reversal potential was determined as –49.5 mV. The experiment was performed with 15 mM NaCl on the *trans* side and 150 mM on the *cis* side.

pared at LPRs of 100 and 200, but not with those prepared at LPRs of 400 and 800. This indicated that the observed currents were due to the incorporated protein.

While the observed currents were typically macroscopic, sometimes smaller currents could be observed that contained transitions characteristic of single channels (Figure 53). However, those transitions corresponded to a conductance of 210 pS, which is large compared to other ion channels which typically have conductances below 100 pS [1].

Similar single channel currents were also observed when different salt concentrations were used on the two sides of the bilayer. Those currents also appeared to have a negative reversal potential, indicating that they were cation selective, which was in agreement with the observed macroscopic currents. For one experiment the reversal potential was determined as –49.5 mV (Figure 54), corresponding to P(Cl)/P(Na) = 0.04.

Since the recorded currents were found to be cation selective, the relative permeability of different cations was also investigated. Currents were observed for different sizes of monovalent and divalent cations but the permeabilities were generally found to increase with the ionic radii (Figure 55). This was consistent with findings from cBEST1 and kpBEST, which were both

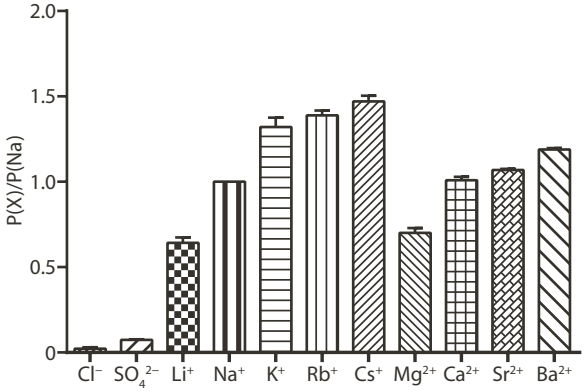


Figure 55 – Selectivity for cations of different sizes. The permeabilities of different ions, compared to that of sodium, are shown as bars. For reference, a bar is also shown for sodium.

more permeable to larger ions [38,67]. kpBEST was, however, not reported to be permeable to divalent cations, unlike the currents described here.

5.1.3. Discussion

The described currents were initially thought to be produced by 33SSP for the following reasons: 1) The protein reconstituted into the liposomes had been carefully purified, thus minimising the risk of contaminations from other proteins. 2) Macroscopic currents were observed, but not from liposomes reconstituted with very small amounts of protein. 3) The currents were cation selective, in agreement with results for kpBEST. 4) Larger permeabilities were observed for larger cations, consistent with studies of kpBEST and bestrophins from animals.

Several observations were, however, made that made me believe that the observed currents might instead be the result of some artefact. These observations are discussed below.

First, cation selective macroscopic currents could also be produced by adding empty liposomes. Initially, this observation was not made, most likely because I had difficulties to reproduce the observed currents in the beginning. Later, as I gained more experience with bilayers, it became easier to reproduce currents from 33SSP-containing proteoliposomes, but eventually I discovered that similar currents could also be produced with empty liposomes added together with 1–2 M KCl. Cation selective currents from empty liposomes have previously been observed by others when the liposomes were added close to the bilayer [149]. It is possible that those currents arise as an artefact from simultaneous fusion of a large number of vesicles, a phenomenon described by others as 'massive vesicle fusion' [150].

Second, no tested mutations changed the selectivity of the observed currents. Based on the crystal structure of 33SSP, I mutated the residues Asp82 and Glu185, believed to be involved in ion selectivity. As 33SSP is a pentamer, each mutation to a non-charged residue was expected to effectively remove five negative charges. Despite this, the currents observed from those mutants remained equally cation selective.

Third, the channel appeared to be permeable to tetraethylammonium. Such quaternary ammonium compounds are known blockers of different potassium channels [1] as well as cation selective pentameric ligand-gated ion channels [151,152]. It thus seemed surprising that this bulky ion would be able to permeate through 33SSP.

Finally, the currents appeared to be originating from a permanently open channel and they were not affected by the addition of calcium or EGTA. This meant that the channel would either be constitutively open or it would be activated by some unknown ligand that was either co-purified with the protein or present in all the buffers. While these options could not be ruled out, they did not seem very likely.

In summary, the possibility remains that a part of the observed currents might be originating from 33SSP. However, it seems likely that a large part, if not all, of the observed currents instead was caused by some artefact. Further experiments, using other techniques, are therefore required in order to study the function of 33SSP. Those experiments will be complicated by the fact that it is not known how bacterial bestrophins are activated. If 33SSP behaves similar to cBEST1, mutations in the neck region could potentially create a channel that would be active in the absence of a ligand. The triple alanine mutation of the neck residues that caused cBEST1 to be active in the absence of calcium [38] could so far not be produced in 33SSP, but other combinations of mutations remain to be investigated.

5.1.4. Methods

For bilayer recordings, small contaminations by bacterial outer membrane porins can be problematic since porins have very high conductances compared to typical ion channels [153]. To avoid this, 33SSP was expressed with a C-terminal streptavidin binding peptide tag [154] and bound to a streptavidin column. The bound protein was extensively washed with a 0.5M NaCl solution (>50 column volumes) followed by a shorter wash with 1M NaCl (~20 column volumes) and elution with biotin. After cleavage of the tag (using SEC purified HRV 3C protease), the protein was purified by SEC on a Superdex 200 column, followed by anion exchange as a polishing step.

The reconstitution into proteoliposomes was performed essentially as described elsewhere [155]. Liposomes were formed from a 3:1 mixture of *E. coli* polar lipids (Avanti Polar Lipids) and L-α-Phosphatidylcholine from egg yolk (Sigma). Large multilamellar vesicles were formed by sonication followed by three freeze/thaw cycles. These were used to form large unilamellar vesicles by extrusion through a 400 nm polycarbonate filter (Avanti Polar Lipids). The produced liposomes were destabilised by the addition of Triton X-100 (Sigma), while monitoring the optical density at 540 nm. The purified protein was mixed with the liposomes at different lipid to protein ratios (LPR). Typically, ratios ranging from 50 to 200 were used. The detergent was removed by incubation with Bio-Beads SM-2 (Bio-Rad). The produced proteoliposomes were flash-frozen in small aliquots that were stored at −80°C until later use.

Protein incorporation in the liposomes was confirmed by extraction with DDM. Non-solubilised lipids were removed by ultracentrifugation. The supernatant was first analysed by silver stained SDS-PAGE, which showed a band of the expected molecular weight. Next the sample was analysed by FSEC (using tryptophan fluorescence), which showed a monodisperse peak eluting at the volume expected for a pentamer.

The bilayer was made in a horizontal setup, where a small piece of overhead foil served as the partition that separated the *trans* and *cis* chambers. A hole in the partition (diameter of

roughly 100 μm) was made by poking the plastic with a needle and shaving off excess plastic on the other side. For all measurements, a 3:1 mixture of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG) (both from Avanti Polar Lipids) dissolved in decane was used to form the bilayer. To each chamber a silver chloride electrode was connected through a salt bridge. The *trans* chamber was held at ground, while the *cis* chamber was connected to the headstage of a Axopatch 200B patch clamp amplifier (Molecular Devices). The data was filtered through a 1 kHz lowpass Bessel filter built into the patch clamp amplifier and digitised using a Digidata 1322A data acquisition system (Molecular Devices). Recording was done using the pCLAMP 9 software suite (Molecular Devices), with the sample rate set to 10 kHz. The headstage and the bilayer assembly was placed in a grounded metal box, which was standing on a vibration isolation table (Kinetic Systems Inc.).

For each day of measurements, a fresh aliquot of frozen proteoliposomes was thawed and briefly sonicated in a bath sonicator. Fusion was induced by mixing the proteoliposomes with 1–2M KCl and adding the mixture on top of the bilayer. If necessary, additional KCl was added to promote fusion. Anion/cation selectivity was investigated using a 10-fold higher NaCl concentration in the *trans* chamber compared to the *cis* chamber (the amount of Na^+ ions in the buffer was taken into account). Voltage clamp experiments were used to determine the reversal potential and the relative permeabilities for Na^+ and Cl^- ions were calculated using the Goldman-Hodgkin-Katz voltage equation [156]. Relative permeabilities of different cations were investigated by using buffers containing 200 mM NaCl in the *cis* chamber and 200 mM of another salt in the *trans* chamber (100 mM was used for divalent cations). The Goldman-Hodgkin-Katz voltage equation was used to calculate relative permeabilities when only monovalent ions were present. When divalent ions were included, a modified form of that equation was used [157].

6. Curriculum Vitae

Name: HOFER (Premarital name: BAGER)

First name: René

Date of birth: 19/12/1983

Nationality: Denmark

Higher education

2011–present Employment as a PhD student at University of Zurich (since 1/8/2011)

2007–2010 M.Sc. in Molecular Biology, Faculty of Science, Aarhus University, Denmark

Graduation date: 18/5/2010

Title of master's thesis: Plasminogen Activation in Zebrafish

2004–2007 B.Sc. in Molecular Biology, Faculty of Science, Aarhus University, Denmark

Graduation date: 8/8/2007

Secondary education

2000–2003 Randers Statsskole, Randers, Denmark

Type of education: "Studentereksamen"

7. Bibliography

1 Hille, B. (2001) Ion channels of excitable membranes, Sinauer, Sunderland, Mass., Great Britain

2 Yu, F. H. and Catterall, W. A. (2003) Overview of the voltage-gated sodium channel family. Genome biology 4, 207

3 Booth, I. R. and Blount, P. (2012) The MscS and MscL families of mechanosensitive channels act as microbial emergency release valves. Journal of bacteriology 194, 4802–4809

4 Anishkin, A., Loukin, S. H., Teng, J. and Kung, C. (2014) Feeling the hidden mechanical forces in lipid bilayer is an original sense. Proceedings of the National Academy of Sciences of the United States of America 111, 7898–7905

5 Jentsch, T. J. (2016) VRACs and other ion channels and transporters in the regulation of cell volume and beyond. Nature reviews. Molecular cell biology 17, 293–307

6 Voets, T., Droogmans, G., Raskin, G., Eggermont, J. and Nilius, B. (1999) Reduced intracellular ionic strength as the initial trigger for activation of endothelial volume-regulated anion channels. Proceedings of the National Academy of Sciences of the United States of America 96, 5298–5303

7 Syeda, R., Qiu, Z., Dubin, A. E., Murthy, S. E., Florendo, M. N., Mason, D. E., Mathur, J., Cahalan, S. M., Peters, E. C. and Montal, M. et al. (2016) LRRC8 Proteins Form Volume-Regulated Anion Channels that Sense Ionic Strength. Cell 164, 499–511

8 Thompson, A. J., Lester, H. A. and Lummis, S. C. R. (2010) The structural basis of function in Cys-loop receptors. Quarterly reviews of biophysics 43, 449–499

9 Vergara, C., Latorre, R., Marrion, N. V. and Adelman, J. P. (1998) Calcium-activated potassium channels. Current opinion in neurobiology 8, 321–329

10 Cui, J., Yang, H. and Lee, U. S. (2009) Molecular mechanisms of BK channel activation. Cellular and molecular life sciences : CMLS 66, 852–875

11 Dascal, N. (2001) Ion-channel regulation by G proteins. Trends in endocrinology and metabolism: TEM 12, 391–398

12 Holzer, P. (2009) Acid-sensitive ion channels and receptors. Handbook of experimental pharmacology, 283–332

13 Bader, C. R., Bertrand, D. and Schwartz, E. A. (1982) Voltage-activated and calcium-activated currents studied in solitary rod inner segments from the salamander retina. The Journal of physiology 331, 253–284

14 Miledi, R. (1982) A calcium-dependent transient outward current in Xenopus laevis oocytes. Proceedings of the Royal Society of London. Series B, Biological sciences 215, 491–497

15 Marty, A., Tan, Y. P. and Trautmann, A. (1984) Three types of calcium-dependent channel in rat lacrimal glands. The Journal of physiology 357, 293–325

16 Barish, M. E. (1983) A transient calcium-dependent chloride current in the immature Xenopus oocyte. The Journal of physiology 342, 309–325

17 Evans, M. G. and Marty, A. (1986) Calcium-dependent chloride currents in isolated cells from rat lacrimal glands. The Journal of physiology 378, 437–460

18 Hartzell, C., Putzier, I. and Arreola, J. (2005) Calcium-activated chloride channels. Annual review of physiology 67, 719–758

19 Arreola, J., Melvin, J. E. and Begenisich, T. (1996) Activation of calcium-dependent chloride channels in rat parotid acinar cells. The Journal of general physiology 108, 35–47

20 Petrukhin, K., Koisti, M. J., Bakall, B., Li, W., Xie, G., Marknell, T., Sandgren, O., Forsman, K., Holmgren, G. and Andreasson, S. et al. (1998) Identification of the gene responsible for Best macular dystrophy. Nature genetics 19, 241–247

21 Stohr, H., Marquardt, A., Nanda, I., Schmid, M. and Weber, B. H. F. (2002) Three novel human VMD2-like genes are members of the evolutionary highly conserved RFP-TM family. European journal of human genetics : EJHG 10, 281–284

22 Sun, H., Tsunenari, T., Yau, K.-W. and Nathans, J. (2002) The vitelliform macular dystrophy protein defines a new family of chloride channels. Proceedings of the National Academy of Sciences of the United States of America 99, 4008–4013

23 Tsunenari, T., Sun, H., Williams, J., Cahill, H., Smallwood, P., Yau, K.-W. and Nathans, J. (2003) Structure-function analysis of the bestrophin family of anion channels. The Journal of biological chemistry 278, 41114–41125

24 Yang, Y. D., Cho, H., Koo, J. Y., Tak, M. H., Cho, Y., Shim, W.-S., Park, S. P., Lee, J., Lee, B. and Kim, B.-M. et al. (2008) TMEM16A confers receptor-activated calcium-dependent chloride conductance. Nature 455, 1210–1215

25 Schroeder, B. C., Cheng, T., Jan, Y. N. and Jan, L. Y. (2008) Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. Cell 134, 1019–1029

26 Caputo, A., Caci, E., Ferrera, L., Pedemonte, N., Barsanti, C., Sondo, E., Pfeffer, U., Ravazzolo, R., Zegarra-Moran, O. and Galietta, L. J. V. (2008) TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. Science (New York, N.Y.) 322, 590–594

27 Pedemonte, N. and Galietta, L. J. V. (2014) Structure and function of TMEM16 proteins (anoctamins). Physiological reviews 94, 419–459

28 Hartzell, H. C., Qu, Z., Yu, K., Xiao, Q. and Chien, L.-T. (2008) Molecular physiology of bestrophins: multifunctional membrane proteins linked to best disease and other retinopathies. Physiological reviews 88, 639–672

29 Hagen, A. R., Barabote, R. D. and Saier, M. H. (2005) The bestrophin family of anion channels: identification of prokaryotic homologues. Molecular membrane biology 22, 291–302

30 Ishida, T. and Kinoshita, K. (2008) Prediction of disordered regions in proteins based on the meta approach. Bioinformatics (Oxford, England) 24, 1344–1348

31 Minezaki, Y., Homma, K. and Nishikawa, K. (2007) Intrinsically disordered regions of human plasma membrane proteins preferentially occur in the cytoplasmic segment. Journal of molecular biology 368, 902–913

32 Milenkovic, V. M., Rivera, A., Horling, F. and Weber, B. H. F. (2007) Insertion and topology of normal and mutant bestrophin-1 in the endoplasmic reticulum membrane. The Journal of biological chemistry 282, 1313–1321

33 Stanton, J. B., Goldberg, A. F. X., Hoppe, G., Marmorstein, L. Y. and Marmorstein, A. D. (2006) Hydrodynamic properties of porcine bestrophin-1 in Triton X-100. Biochimica et biophysica acta 1758, 241–247

34 Qu, Z. and Hartzell, H. C. (2008) Bestrophin Cl- channels are highly permeable to HCO3-. American journal of physiology. Cell physiology 294, 7

35 O'Driscoll, K. E., Leblanc, N., Hatton, W. J. and Britton, F. C. (2009) Functional properties of murine bestrophin 1 channel. Biochemical and biophysical research communications 384, 476–481

36 Park, H., Oh, S.-J., Han, K.-S., Woo, D. H., Park, H., Mannaioni, G., Traynelis, S. F. and Lee, C. J. (2009) Bestrophin-1 encodes for the Ca2+-activated anion channel in hippocampal astrocytes. The Journal of neuroscience : the official journal of the Society for Neuroscience 29, 13063–13073

37 Kane Dickson, V., Pedi, L. and Long, S. B. (2014) Structure and insights into the function of a Ca(2+)-activated Cl(-) channel. Nature 516, 213–218

38 Vaisey, G., Miller, A. N. and Long, S. B. (2016) Distinct regions that control ion selectivity and calcium-dependent activation in the bestrophin ion channel. Proceedings of the National Academy of Sciences of the United States of America 113, E7399-E7408

39 Qu, Z., Fischmeister, R. and Hartzell, C. (2004) Mouse bestrophin-2 is a bona fide Cl(-) channel: identification of a residue important in anion binding and conduction. The Journal of general physiology 123, 327–340

40 Tsunenari, T., Nathans, J. and Yau, K.-W. (2006) Ca2+-activated Cl- current from human bestrophin-4 in excised membrane patches. The Journal of general physiology 127, 749–754

41 Xiao, Q., Prussia, A., Yu, K., Cui, Y.-y. and Hartzell, H. C. (2008) Regulation of bestrophin Cl channels by calcium: role of the C terminus. The Journal of general physiology 132, 681–692

42 Yuan, P., Leonetti, M. D., Pico, A. R., Hsiung, Y. and MacKinnon, R. (2010) Structure of the human BK channel Ca2+-activation apparatus at 3.0 A resolution. Science (New York, N.Y.) 329, 182–186

43 Marmorstein, L. Y., McLaughlin, P. J., Stanton, J. B., Yan, L., Crabb, J. W. and Marmorstein, A. D. (2002) Bestrophin interacts physically and functionally with protein phosphatase 2A. The Journal of biological chemistry 277, 30591–30597

44 Xiao, Q., Yu, K., Cui, Y.-y. and Hartzell, H. C. (2009) Dysregulation of human bestrophin-1 by ceramide-induced dephosphorylation. The Journal of physiology 587, 4379–4391

45 Duran, C., Chien, L.-T. and Hartzell, H. C. (2013) Drosophila bestrophin-1 currents are regulated by phosphorylation via a CaMKII dependent mechanism. PloS one 8, e58875

46 Qu, Z., Cui, Y. and Hartzell, C. (2006) A short motif in the C-terminus of mouse bestrophin 3 corrected inhibits its activation as a Cl channel. FEBS letters 580, 2141–2146

47 Qu, Z. Q., Yu, K., Cui, Y. Y., Ying, C. and Hartzell, C. (2007) Activation of bestrophin Cl-channels is regulated by C-terminal domains. The Journal of biological chemistry 282, 17460–17467

48 Chien, L.-T. and Hartzell, H. C. (2007) Drosophila bestrophin-1 chloride current is dually regulated by calcium and cell volume. The Journal of general physiology 130, 513–524

49 Stotz, S. C. and Clapham, D. E. (2012) Anion-sensitive fluorophore identifies the Drosophila swell-activated chloride channel in a genome-wide RNA interference screen. PloS one 7, e46865

50 Chien, L.-T. and Hartzell, H. C. (2008) Rescue of volume-regulated anion current by bestrophin mutants with altered charge selectivity. The Journal of general physiology 132, 537–546

51 Qiu, Z., Dubin, A. E., Mathur, J., Tu, B., Reddy, K., Miraglia, L. J., Reinhardt, J., Orth, A. P. and Patapoutian, A. (2014) SWELL1, a plasma membrane protein, is an essential component of volume-regulated anion channel. Cell 157, 447–458

52 Milenkovic, A., Brandl, C., Milenkovic, V. M., Jendryke, T., Sirianant, L., Wanitchakool, P., Zimmermann, S., Reiff, C. M., Horling, F. and Schrewe, H. et al. (2015) Bestrophin 1 is indispensable for volume regulation in human retinal pigment epithelium cells. Proceedings of the National Academy of Sciences of the United States of America 112, 9

53 Fischmeister, R. and Hartzell, H. C. (2005) Volume sensitivity of the bestrophin family of chloride channels. The Journal of physiology 562, 477–491

54 Rosenthal, R., Bakall, B., Kinnick, T., Peachey, N., Wimmers, S., Wadelius, C., Marmorstein, A. and Strauss, O. (2006) Expression of bestrophin-1, the product of the VMD2 gene, modulates voltage-dependent Ca2+ channels in retinal pigment epithelial cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 20, 178–180

55 Reichhart, N., Milenkovic, V. M., Halsband, C.-A., Cordeiro, S. and Strauss, O. (2010) Effect of bestrophin-1 on L-type Ca2+ channel activity depends on the Ca2+ channel beta-subunit. Experimental eye research 91, 630–639

56 Milenkovic, V. M., Krejcova, S., Reichhart, N., Wagner, A. and Strauss, O. (2011) Interaction of bestrophin-1 and Ca2+ channel beta-subunits: identification of new binding domains on the bestrophin-1 C-terminus. PloS one 6, e19364

57 Yu, K., Xiao, Q., Cui, G., Lee, A. and Hartzell, H. C. (2008) The best disease-linked Cl- channel hBest1 regulates Ca V 1 (L-type) Ca2+ channels via src-homology-binding domains. The Journal of neuroscience : the official journal of the Society for Neuroscience 28, 5660–5670

58 Park, H., Han, K.-S., Oh, S.-J., Jo, S., Woo, J., Yoon, B.-E. and Lee, C. J. (2013) High glutamate permeability and distal localization of Best1 channel in CA1 hippocampal astrocyte. Molecular brain 6, 54

59 Lee, S., Yoon, B.-E., Berglund, K., Oh, S.-J., Park, H., Shin, H.-S., Augustine, G. J. and Lee, C. J. (2010) Channel-mediated tonic GABA release from glia. Science (New York, N.Y.) 330, 790–796

60 Woo, D. H., Han, K.-S., Shim, J. W., Yoon, B.-E., Kim, E., Bae, J. Y., Oh, S.-J., Hwang, E. M., Marmorstein, A. D. and Bae, Y. C. et al. (2012) TREK-1 and Best1 channels mediate fast and slow glutamate release in astrocytes upon GPCR activation. Cell 151, 25–40

61 Tochitani, S. and Kondo, S. (2013) Immunoreactivity for GABA, GAD65, GAD67 and Bestrophin-1 in the meninges and the choroid plexus: implications for non-neuronal sources for GABA in the developing mouse brain. PloS one 8, e56901

62 Marmorstein, A. D., Cross, H. E. and Peachey, N. S. (2009) Functional roles of bestrophins in ocular epithelia. Progress in retinal and eye research 28, 206–226

63 Yu, K., Lujan, R., Marmorstein, A., Gabriel, S. and Hartzell, H. C. (2010) Bestrophin-2 mediates bicarbonate transport by goblet cells in mouse colon. The Journal of clinical investigation 120, 1722–1735

64 Cui, C.-Y., Childress, V., Piao, Y., Michel, M., Johnson, A. A., Kunisada, M., Ko, M. S. H., Kaestner, K. H., Marmorstein, A. D. and Schlessinger, D. (2012) Forkhead transcription factor FoxA1 regulates sweat secretion through Bestrophin 2 anion channel and Na-K-Cl cotransporter 1. Proceedings of the National Academy of Sciences of the United States of America 109, 1199–1203

65 Pifferi, S., Pascarella, G., Boccaccio, A., Mazzatenta, A., Gustincich, S., Menini, A. and Zucchelli, S. (2006) Bestrophin-2 is a candidate calcium-activated chloride channel involved in olfactory transduction. Proceedings of the National Academy of Sciences of the United States of America 103, 12929–12934

66 Roberts, S. K., Milnes, J. and Caddick, M. (2011) Characterisation of AnBEST1, a functional anion channel in the plasma membrane of the filamentous fungus, Aspergillus nidulans. Fungal genetics and biology : FG & B 48, 928–938

67 Yang, T., Liu, Q., Kloss, B., Bruni, R., Kalathur, R. C., Guo, Y., Kloppmann, E., Rost, B., Colecraft, H. M. and Hendrickson, W. A. (2014) Structure and selectivity in bestrophin ion channels. Science (New York, N.Y.) 346, 355–359

68 Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M. and Ostell, J. (2016) NCBI prokaryotic genome annotation pipeline. Nucleic acids research 44, 6614–6624

69 Foster, J. W. and Davis, R. H. (1966) A methane-dependent coccus, with notes on classification and nomenclature of obligate, methane-utilizing bacteria. Journal of bacteriology 91, 1924–1931

70 Yamaoka, T., Satoh, K. and Katoh, S. (1978) Photosynthetic activities of a thermophilic blue-green alga. Plant and Cell Physiology 19, 943–954

71 Kawate, T. and Gouaux, E. (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure (London, England : 1993) 14, 673–681

72 Sonoda, Y., Cameron, A., Newstead, S., Omote, H., Moriyama, Y., Kasahara, M., Iwata, S. and Drew, D. (2010) Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins. FEBS letters 584, 2539–2547

73 Waldo, G. S., Standish, B. M., Berendzen, J. and Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein. Nature biotechnology 17, 691–695

74 Geertsma, E. R., Groeneveld, M., Slotboom, D.-J. and Poolman, B. (2008) Quality control of overexpressed membrane proteins. Proceedings of the National Academy of Sciences of the United States of America 105, 5722–5727

75 Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. Journal of bacteriology 177, 4121–4130

76 Geertsma, E. R., Chang, Y.-N., Shaik, F. R., Neldner, Y., Pardon, E., Steyaert, J. and Dutzler, R. (2015) Structure of a prokaryotic fumarate transporter reveals the architecture of the SLC26 family. Nature structural & molecular biology 22, 803–808

77 Ehrnstorfer, I. A., Geertsma, E. R., Pardon, E., Steyaert, J. and Dutzler, R. (2014) Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport. Nature structural & molecular biology 21, 990–996

78 Ostermeier, C. and Michel, H. (1997) Crystallization of membrane proteins. Current opinion in structural biology 7, 697–701

79 Sonoda, Y., Newstead, S., Hu, N.-J., Alguel, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J. and Cameron, A. D. et al. (2011) Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. Structure (London, England : 1993) 19, 17–25

80 Hattori, M., Hibbs, R. E. and Gouaux, E. (2012) A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. Structure (London, England : 1993) 20, 1293–1299

81 Mancusso, R., Karpowich, N. K., Czyzewski, B. K. and Wang, D.-N. (2011) Simple screening method for improving membrane protein thermostability. Methods (San Diego, Calif.) 55, 324–329

82 Gourdon, P., Andersen, J. L., Hein, K. L., Bublitz, M., Pedersen, B. P., Liu, X.-Y., Yatime, L., Nyblom, M., Nielsen, T. T. and Olesen, C. et al. (2011) HiLiDe—Systematic Approach to Membrane Protein Crystallization in Lipid and Detergent. Crystal Growth & Design 11, 2098–2106

83 Ostermeier, C., Iwata, S., Ludwig, B. and Michel, H. (1995) Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. Nature structural biology 2, 842–846

84 Hunte, C. and Michel, H. (2002) Crystallisation of membrane proteins mediated by antibody fragments. Current opinion in structural biology 12, 503–508

85 Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. and Koide, S. (2007) High-affinity single-domain binding proteins with a binary-code interface. Proceedings of the National Academy of Sciences of the United States of America 104, 6632–6637

86 Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. and Pluckthun, A. (2003) Designing repeat proteins. Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. Journal of molecular biology 332, 489–503

87 Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N. and Hamers, R. (1993) Naturally occurring antibodies devoid of light chains. Nature 363, 446–448

88 Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C. and Flajnik, M. F. (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374, 168–173

89 Arbabi Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. and Muyldermans, S. (1997) Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. FEBS letters 414, 521–526

90 Genst, E. de, Silence, K., Decanniere, K., Conrath, K., Loris, R., Kinne, J., Muyldermans, S. and Wyns, L. (2006) Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proceedings of the National Academy of Sciences of the United States of America 103, 4586–4591

91 Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E. and Calinski, D. et al. (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature 477, 549–555

92 Beckett, D., Kovaleva, E. and Schatz, P. J. (1999) A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. Protein science : a publication of the Protein Society 8, 921–929

93 Harmsen, M. M., Ruuls, R. C., Nijman, I. J., Niewold, T. A., Frenken, L. G. and Geus, B. de (2000) Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. Molecular immunology 37, 579–590

94 Muyldermans, S. (2013) Nanobodies. Natural single-domain antibodies. Annual review of biochemistry 82, 775–797

95 Salema, V. and Fernandez, L. A. (2013) High yield purification of nanobodies from the periplasm of E. coli as fusions with the maltose binding protein. Protein expression and purification 91, 42–48

96 Karplus, P. A. and Diederichs, K. (2015) Assessing and maximizing data quality in macromolecular crystallography. Current opinion in structural biology 34, 60–68

97 Mueller, M., Wang, M. and Schulze-Briese, C. (2012) Optimal fine phi-slicing for single-photon-counting pixel detectors. Acta crystallographica. Section D, Biological crystallography 68, 42–56

98 Kabsch, W. (2010) XDS. Acta crystallographica. Section D, Biological crystallography 66, 125–132

99 Evans, P. (2006) Scaling and assessment of data quality. Acta crystallographica. Section D, Biological crystallography 62, 72–82

100 Karplus, P. A. and Diederichs, K. (2012) Linking crystallographic model and data quality. Science (New York, N.Y.) 336, 1030–1033

101 Kantardjieff, K. A. and Rupp, B. (2003) Matthews coefficient probabilities. Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein science : a publication of the Protein Society 12, 1865–1871

102 Matthews, B. W. (1968) Solvent content of protein crystals. Journal of molecular biology 33, 491–497

103 Evans, P. and McCoy, A. (2008) An introduction to molecular replacement. Acta crystallographica. Section D, Biological crystallography 64, 1–10

104 Scapin, G. (2013) Molecular replacement then and now. Acta crystallographica. Section D, Biological crystallography 69, 2266–2275

105 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. Journal of applied crystallography 40, 658–674

106 Bunkoczi, G. and Read, R. J. (2011) Improvement of molecular-replacement models with Sculptor. Acta crystallographica. Section D, Biological crystallography 67, 303–312

107 Dodson, E. (2008) The befores and afters of molecular replacement. Acta crystallographica. Section D, Biological crystallography 64, 17–24

108 Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. Acta crystallographica. Section D, Biological crystallography 68, 352–367

109 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J. and Grosse-Kunstleve, R. W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica. Section D, Biological crystallography 66, 213–221

110 Terwilliger, T. C., Read, R. J., Adams, P. D., Brunger, A. T., Afonine, P. V. and Hung, L.-W. (2013) Model morphing and sequence assignment after molecular replacement. Acta crystallographica. Section D, Biological crystallography 69, 2244–2250

111 Terwilliger, T. C. (2004) Using prime-and-switch phasing to reduce model bias in molecular replacement. Acta crystallographica. Section D, Biological crystallography 60, 2144–2149

112 Vagin, A. and Teplyakov, A. (2010) Molecular replacement with MOLREP. Acta crystallographica. Section D, Biological crystallography 66, 22–25

113 Headd, J. J., Echols, N., Afonine, P. V., Moriarty, N. W., Gildea, R. J. and Adams, P. D. (2014) Flexible torsion-angle noncrystallographic symmetry restraints for improved macromolecular structure refinement. Acta crystallographica. Section D, Biological crystallography 70, 1346–1356

114 Painter, J. and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta crystallographica. Section D, Biological crystallography 62, 439–450

115 Marmorstein, A. D., Marmorstein, L. Y., Rayborn, M., Wang, X., Hollyfield, J. G. and Petrukhin, K. (2000) Bestrophin, the product of the Best vitelliform macular dystrophy gene (VMD2), localizes to the basolateral plasma membrane of the retinal pigment epithelium. Proceedings of the National Academy of Sciences of the United States of America 97, 12758–12763

116 Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. and Lomize, A. L. (2012) OPM database and PPM web server. Resources for positioning of proteins in membranes. Nucleic acids research 40, D370-6

117 Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. Journal of molecular biology 372, 774–797

118 Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta crystallographica. Section D, Biological crystallography 60, 2256–2268

119 Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) Features and development of Coot. Acta crystallographica. Section D, Biological crystallography 66, 486–501

120 Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. and Sansom, M. S. (1996) HOLE. A program for the analysis of the pore dimensions of ion channel structural models. Journal of molecular graphics 14, 354-60, 376

121 Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R., Muyldermans, S. and Wyns, L. (1996) Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nature structural biology 3, 803–811

122 Spinelli, S., Frenken, L., Bourgeois, D., Ron, L. de, Bos, W., Verrips, T., Anguille, C., Cambillau, C. and Tegoni, M. (1996) The crystal structure of a llama heavy chain variable domain. Nature structural biology 3, 752–757

123 Bork, P., Holm, L. and Sander, C. (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. Journal of molecular biology 242, 309–320

124 Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure. Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637

125 Shannon, R. D. (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. Acta Cryst A 32, 751–767

126 Aryal, P., Sansom, M. S. P. and Tucker, S. J. (2015) Hydrophobic gating in ion channels. Journal of molecular biology 427, 121–130

127 Hibbs, R. E. and Gouaux, E. (2011) Principles of activation and permeation in an anion-selective Cys-loop receptor. Nature 474, 54–60

128 Keramidas, A., Moorhouse, A. J., Schofield, P. R. and Barry, P. H. (2004) Ligand-gated ion channels. Mechanisms underlying ion selectivity. Progress in biophysics and molecular biology 86, 161–204

129 Imoto, K., Busch, C., Sakmann, B., Mishina, M., Konno, T., Nakai, J., Bujo, H., Mori, Y., Fukuda, K. and Numa, S. (1988) Rings of negatively charged amino acids determine the acetylcholine receptor channel conductance. Nature 335, 645–648

130 Konno, T., Busch, C., Kitzing, E. von, Imoto, K., Wang, F., Nakai, J., Mishina, M., Numa, S. and Sakmann, B. (1991) Rings of anionic amino acids as structural determinants of ion selectivity in the acetylcholine receptor channel. Proceedings. Biological sciences 244, 69–79

131 Hilf, R. J. C. and Dutzler, R. (2009) Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. Nature 457, 115–118

132 Corringer, P. J., Bertrand, S., Galzi, J. L., Devillers-Thiery, A., Changeux, J. P. and Bertrand, D. (1999) Mutational analysis of the charge selectivity filter of the alpha7 nicotinic acetylcholine receptor. Neuron 22, 831–843

133 Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R. and Magrane, M. et al. (2004) UniProt. The Universal Protein knowledgebase. Nucleic acids research 32, D115-9

134 CLC Main Workbench, https://www.qiagenbioinformatics.com/

135 Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. and Soding, J. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology 7, 539

136 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59, 307–321

137 Pei, J., Kim, B.-H. and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research 36, 2295–2300

138 Katoh, K. and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7. Improvements in performance and usability. Molecular biology and evolution 30, 772–780

139 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics (Oxford, England) 25, 1189–1191

140 Geertsma, E. R. and Dutzler, R. (2011) A versatile and efficient high-throughput cloning tool for structural biology. Biochemistry 50, 3272–3278

141 Geertsma, E. R., Duurkens, R. H. and Poolman, B. (2005) The activity of the lactose transporter from Streptococcus thermophilus is increased by phosphorylated IIA and the action of beta-galactosidase. Biochemistry 44, 15889–15897

142 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In The Proteomics Protocols Handbook (Walker, J. M., ed.), pp. 571–607

143 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W. and McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. Acta crystallographica. Section D, Biological crystallography 67, 235–242

144 Vincke, C., Gutierrez, C., Wernery, U., Devoogdt, N., Hassanzadeh-Ghassabeh, G. and Muyldermans, S. (2012) Generation of single domain antibody fragments derived from

camelids and generation of manifold constructs. Methods in molecular biology (Clifton, N.J.) 907, 145–176

145 Hassanzadeh-Ghassabeh, G., Saerens, D. and Muyldermans, S. (2011) Generation of anti-infectome/anti-proteome nanobodies. Methods in molecular biology (Clifton, N.J.) 790, 239–259

146 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry 25, 1605–1612

147 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) Electrostatics of nanosystems. Application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America 98, 10037–10041

148 Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. and Baker, N. A. (2004) PDB2PQR. An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic acids research 32, W665-7

149 Woodbury, D. J. (1989) Pure lipid vesicles can induce channel-like conductances in planar bilayers. The Journal of membrane biology 109, 145–150

150 Labarca, P. and Latorre, R. (1992) [30] Insertion of ion channels into planar lipid bilayers by vesicle fusion. Methods in enzymology 207, 447–463

151 Adler, M., Oliveira, A. C., Albuquerque, E. X., Mansour, N. A. and Eldefrawi, A. T. (1979) Reaction of tetraethylammonium with the open and closed conformations of the acetylcholine receptor ionic channel complex. The Journal of general physiology 74, 129

152 Hilf, R. J. C., Bertozzi, C., Zimmermann, I., Reiter, A., Trauner, D. and Dutzler, R. (2010) Structural basis of open channel block in a prokaryotic pentameric ligand-gated ion channel. Nature structural & molecular biology 17, 1330–1336

153 Accardi, A. and Miller, C. (2004) Secondary active transport mediated by a prokaryotic homologue of ClC Cl- channels. Nature 427, 803–807

154 Keefe, A. D., Wilson, D. S., Seelig, B. and Szostak, J. W. (2001) One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. Protein expression and purification 23, 440–446

155 Geertsma, E. R., Nik Mahmood, N A B, Schuurman-Wolters, G. K. and Poolman, B. (2008) Membrane reconstitution of ABC transporters and assays of translocator function. Nature protocols 3, 256–266

156 Hodgkin, A. L. and Katz, B. (1949) The effect of sodium ions on the electrical activity of giant axon of the squid. The Journal of physiology 108, 37–77

157 Lewis, C. A. (1979) Ion-concentration dependence of the reversal potential and the single channel conductance of ion channels at the frog neuromuscular junction. The Journal of physiology 286, 417–445